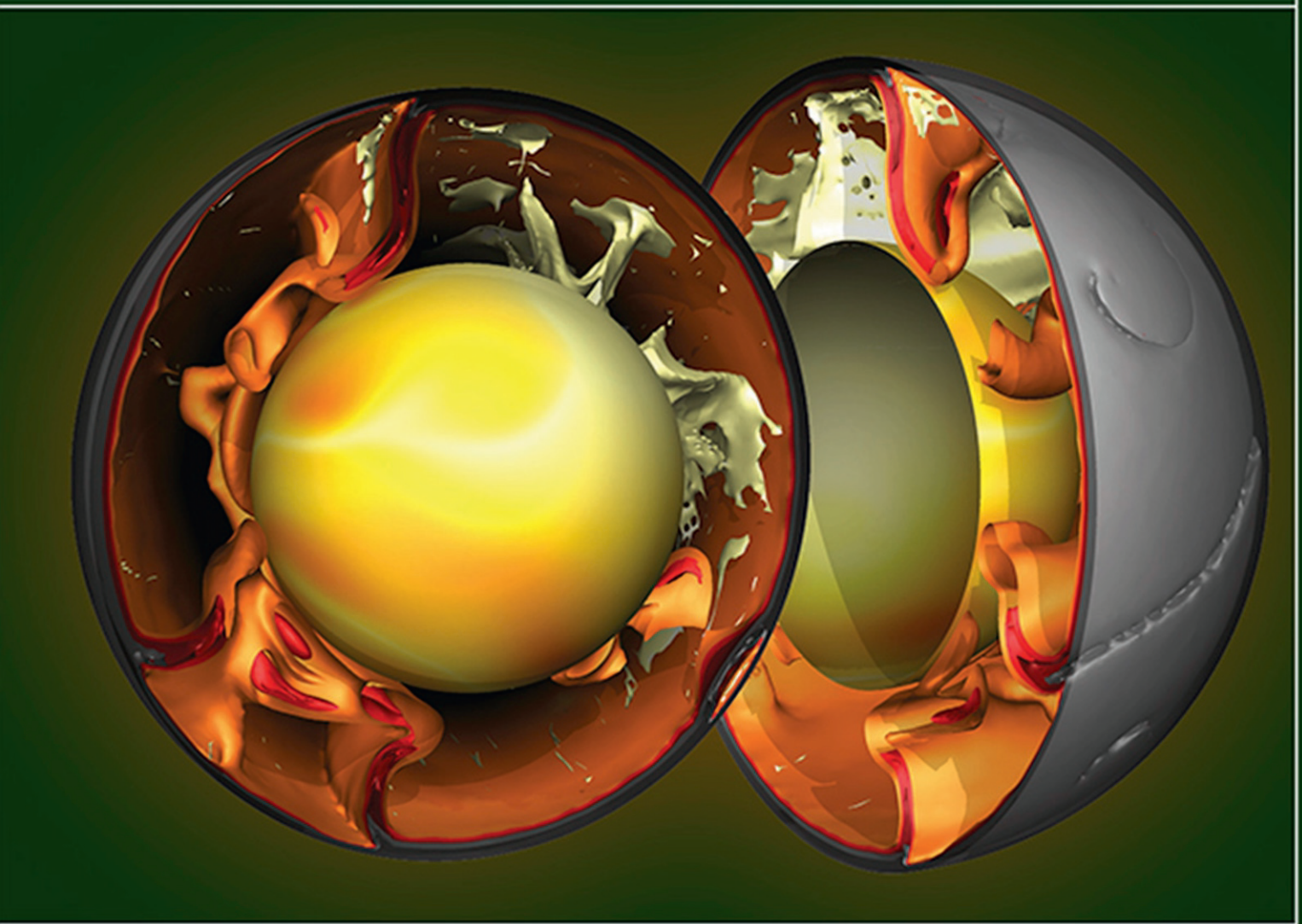
SECOND EDITION

TREATISE ON GEOPHYSICS

Volume 7

MANTLE DYNAMICS



VOLUME EDITOR: DAVID BERCOVICI

EDITOR-IN-CHIEF: GERALD SCHUBERT

TREATISE ON GEOPHYSICS

SECOND EDITION

This page intentionally left blank

TREATISE ON GEOPHYSICS

SECOND EDITION

EDITOR-IN-CHIEF

Gerald Schubert

Distinguished Professor Emeritus, Department of Earth, Planetary and Space Sciences, University of California, Los Angeles

VOLUME 7

MANTLE DYNAMICS

VOLUME EDITOR

Professor David Bercovici

Department of Geology and Geophysics, Yale University

ELSEVIER AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
30 Corporate Drive, Suite 400, Burlington MA 01803, USA

Copyright © 2015 Elsevier B.V. unless otherwise stated. All rights reserved.

Cover image created for this work by Fabio Crameri, University College London.
The figure depicts the temperature field from a convection calculation using the code STAG-YY, written by Paul Tackley.

The following chapters are US Government works in the public domain and not subject to copyright:

1.11 Crust and Lithospheric Structure - Global Crustal Structure; 3.09 Earth Rotation Variations – Long Period; 4.07 The Role of Fault-Zone Drilling; 5.11 The Time-Averaged Field and Paleosecular Variation; 10.21 Exoplanetary Geophysics: An Emerging Discipline; 11.12 Tools and Techniques: Gravitational Method.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN (print): 978-0-444-53802-4

For information on all Elsevier publications
visit our website at http://store.elsevier.com/

Editorial & Production: Gemma Tomalin, Richard Berryman, Marise Willis, Michael Nicholls, Ellen Bishop
Cover design: Matthew Limbert

Printed and bound in the United Kingdom

15 16 17 18 19 10 9 8 7 6 5 4 3 2 1

CONTENTS

Preface *vii*

Editor-in-Chief *ix*

Contributors *xi*

Contents of all Volumes *xiii*

7.01 Mantle Dynamics: An Introduction and Overview 1
D Bercovici

7.02 Physics of Mantle Convection 23
Y Ricard

7.03 Laboratory Studies of Mantle Convection 73
A Davaille and A Limare

7.04 Analytical Approaches to Mantle Dynamics 145
Neil M Ribe

7.05 Numerical Methods for Mantle Convection 197
SJ Zhong, DA Yuen, LN Moresi, and MG Knepley

7.06 Temperatures, Heat, and Energy in the Mantle of the Earth 223
C Jaupart, S Labrosse, F Lucazeau, and J-C Mareschal

7.07 The Generation of Plate Tectonics from Mantle Dynamics 271
D Bercovici, P Tackley, and Y Ricard

7.08 The Dynamics and Convective Evolution of the Upper Mantle 319
EM Parmentier

7.09 Dynamics of Subducting Slabs: Numerical Modeling and Constraints from Seismology, Geoid, Topography, Geochemistry, and Petrology 339
I Wada and S King

7.10 Hotspots, Large Igneous Provinces, and Melting Anomalies 393
MD Ballmer, PE van Keken, and G Ito

7.11 The Core–Mantle Boundary Region 461
JW Hernlund and AK McNamara

7.12 Mantle Geochemical Geodynamics 521
PJ Tackley

This page intentionally left blank

PREFACE

The *Treatise on Geophysics* is a comprehensive and in-depth study of the physics of the Earth and of terrestrial-like planets residing both within and outside our solar system. Its breadth and detail of coverage are beyond what any single geophysics text can provide. The first edition of the *Treatise on Geophysics* was published in 2007, nearly a decade ago. Of course, there is much progress in science in that length of time and the field of geophysics has grown rapidly and developed new lines of inquiry. Accordingly, we have brought the *Treatise on Geophysics* up-to-date with this second edition. The new edition will continue to provide students and professionals with fundamental and state-of-the-art discussion of all aspects of geophysics. In this new edition, the reader will find updates to all the chapters contained in the first edition and new chapters that discuss topics missing from the first edition. In a few instances, chapters from the first edition have simply been reprinted in order to avoid gaps in coverage. Chapters that emphasized fundamental physics often required little updating.

A highlight of the second edition is a new volume on Resources in the Near-Surface Earth. The new volume discusses the role of geophysics in the exploitation and conservation of natural resources and the assessment of degradation of natural systems by pollution. The near surface is a zone where humans and natural systems interact. Understanding the effects of human impacts is a challenge particularly for the long term.

There are four new chapters in the volume on Deep Earth Seismology. One of them describes progress in seismic instrumentation in the last 30 years in the era of broadband digital sensor and recording systems. Another one addresses a new approach for the determination of crust and uppermost mantle structure, based on ambient noise cross-correlations. This has revolutionized the field, since it allows the study of regional structure at high resolution in areas where there are no earthquakes. The third new chapter is devoted to recent advances in characterizing the lithosphere–asthenosphere boundary under both oceans and continents. Finally, the last new chapter discusses the morphology of subducted slabs in the Earth's transition zone.

The Mineral Physics volume has new chapters on dynamic (shock-wave) compression methods, and on the mineralogy of the upper mantle and the D″ layer focusing on post-perovskite, and extrasolar super-Earths.

New material in the Geodesy volume expands the temporal resolution and latency of geodetic measurements to high-rate, real-time measurements and increases the spatial resolution to meter size pixels with optical and radar imaging systems. Application areas are extended with new material on the uses of Earth gravity measurements.

The Earthquake Seismology volume has a new chapter on rock mechanics that is a comprehensive account of modern rock mechanics with emphasis on the variation of frictional properties of rocks during an earthquake cycle. Many chapters have considerable updates reflecting the lessons from significant recent earthquakes (e.g., the 2011 Tohoku-Oki earthquake and the 2013 Okhotsk deep earthquake), new observations on slow slips and triggered earthquakes, and developments of new methods (e.g., back projection method). Some chapters are updated (e.g., paleoseismology and strong ground motion) with substantial new material.

Chapters in the Geomagnetism volume have been updated to include recent advances in the acquisition of larger data sets and higher quality data and in the construction of more reliable and sophisticated models. Improvements in data have occurred in areas such as magnetic anomalies, paleointensities, and planetary magnetism. Improvements in models of the present day field include so-called comprehensive models and models that incorporate data assimilation.

The Crustal and Lithosphere Dynamics volume has new contributions on mountain building and landscape evolution.

The Mantle Dynamics volume contains new chapters on the generation of plate tectonics from mantle convection and the core–mantle boundary region.

The Core Dynamics volume features a comprehensive update of this rapidly changing subject, including new chapters on turbulence and waves in the fluid outer core.

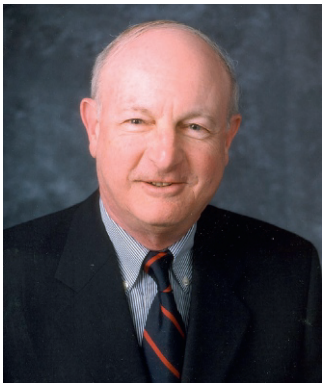
The volume on the Evolution of the Earth presents a new assessment on the role of volatiles in Earth evolution.

There are new chapters in the Physics of Terrestrial Planets and Moons volume on planetary atmospheres, planetary tectonics, extrasolar planets and asteroids and comets. The study of extrasolar planets has had explosive growth in the last decade and new spacecraft missions have significantly enhanced our knowledge of solar system planets and moons.

Finally, this is an opportunity for me to thank and acknowledge the editors of the individual volumes for their hard work, patience, dedication, and commitment in bringing this new edition of the Treatise to fruition. I am also indebted to Gemma Tomalin of Elsevier for her care and attention to the organization and production of the Treatise and to Deepa Arumugam for her careful coordination of the proofs from the copyeditor and typesetter. Of course, the authors of the individual chapters are the ones who are responsible for the excellence of the Treatise. We have all had some ups and downs in the process of producing the second edition of the Treatise but the final product is one that we can all view with pride.

Gerald Schubert
Editor-in-Chief

EDITOR-IN-CHIEF



Gerald Schubert is Distinguished Professor Emeritus in the Department of Earth, Planetary and Space Sciences at the University of California, Los Angeles. His research interests encompass the physics of the interiors and atmospheres of the Earth, the Moon, and the other moons and planets of the solar system. He is co-author with Donald Turcotte of *Geodynamics* (ed. 3, Cambridge University Press, 2014) and with Turcotte and Peter Olson of *Mantle Convection in the Earth and Planets* (Cambridge University Press, 2001). He is the author of over 540 research papers. He has participated in a number of NASA's planetary missions, including Apollo, Pioneer Venus, Magellan, and Galileo, and has been editor and editorial board member of many journals, including *Icarus*, *Journal of Geophysical Research*, *Geophysical Research Letters*, and *Annual Reviews of Earth and Planetary Sciences*. Professor Schubert is a Fellow of the American Geophysical Union and a recipient of the Union's James B. MacElwane medal and the Harry H. Hess medal. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences.

This page intentionally left blank

CONTRIBUTORS

MD Ballmer
University of Hawaii, Honolulu, HI, USA; Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan

D Bercovici
Yale University, New Haven, CT, USA

A Davaille
CNRS/University of Paris-Sud, Orsay, France

JW Hernlund
Tokyo Institute of Technology, Tokyo, Japan

G Ito
University of Hawaii, Honolulu, HI, USA

C Jaupart
Institut de Physique du Globe de Paris, Paris, France

S King
Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

MG Knepley
University of Chicago, Chicago, IL, USA

S Labrosse
Ecole Normale Supérieure de Lyon, Lyon, France

A Limare
Sorbonne Paris Cité/Université Paris-Diderot/CNRS, Paris, France

F Lucazeau
Institut de Physique du Globe de Paris, Paris, France

J-C Mareschal
Université du Québec à Montréal, Montréal, QC, Canada

AK McNamara
Arizona State University, Tempe, AZ, USA

LN Moresi
University of Melbourne, Melbourne, VIC, Australia

EM Parmentier
Brown University, Providence, RI, USA

Neil M Ribe
Université Paris-Sud and CNRS, Orsay, France

Y Ricard
Ecole Normale Supérieure de Lyon, Université de Lyon-1, CNRS, Lyon, France

PJ Tackley
Institut für Geophysik, ETH Zürich, Switzerland

PE van Keken
University of Michigan, Ann Arbor, MI, USA

I Wada
Tohoku University, Sendai, Miyagi, Japan

DA Yuen
University of Minnesota, Minneapolis, MN, USA; China University of Geosciences, Wuhan, China

SJ Zhong
University of Colorado, Boulder, CO, USA

This page intentionally left blank

7.01 Mantle Dynamics: An Introduction and Overview

D Bercovici, Yale University, New Haven, CT, USA

© 2015 Elsevier B.V. All rights reserved.

7.01.1	**Introduction**	1
7.01.2	**A Historical Perspective on Mantle Dynamics**	1
7.01.2.1	Benjamin Thompson, Count Rumford (1753–1814)	2
7.01.2.2	John William Strutt, Lord Rayleigh (1842–1919)	4
7.01.2.3	Henri Claude Bénard (1874–1939)	6
7.01.2.4	Arthur Holmes (1890–1965)	7
7.01.2.5	Anton Hales and Chaim Pekeris	9
7.01.2.5.1	Anton Linder Hales (1911–2006)	9
7.01.2.5.2	Chaim Leib Pekeris (1908–93)	10
7.01.2.6	Harry Hammond Hess (1906–69)	11
7.01.2.7	Stanley Keith Runcorn (1922–95)	13
7.01.2.8	Mantle Convection Theory in the Last 40 Years	14
7.01.3	**Observations and Evidence for Mantle Convection**	14
7.01.4	**Mantle Properties**	15
7.01.5	**Questions About Mantle Convection We Have Probably Answered**	16
7.01.5.1	Does the Mantle Convect?	16
7.01.5.2	Is the Mantle Layered at 660 km?	16
7.01.5.3	What Are the Driving Forces of Tectonic Plates?	16
7.01.6	**Major Unsolved Issues in Mantle Dynamics**	17
7.01.6.1	Energy Sources for Mantle Convection	17
7.01.6.2	Is the Mantle Well Mixed, Layered, Piled, or Plum Pudding?	17
7.01.6.3	Are There Mantle Plumes?	17
7.01.6.4	Origin and Cause of Plate Tectonics	18
7.01.7	**Burgeoning and Future Problems in Mantle Dynamics**	18
7.01.7.1	Volatile Circulation	18
7.01.7.2	Mantle Convection: Water and Life	19
7.01.8	**Summary and Context of the Rest of This Volume**	19
Acknowledgments		21
References		21

7.01.1 Introduction

Much of what we refer to as geology, or more accurately geologic activity on Earth, is due to the simple act of our planet cooling to space. What allows this activity to persist over the lifetime of the solar system is that the most massive portion of the planet, namely, the mantle, is so large, moves so slowly, and cools so gradually that it sets the pace of cooling for the whole Earth. If the Earth's other components, such as the crust and core, were allowed to lose heat on their own, their small size or facile motion would have allowed them to cool rapidly and their activity would have ceased aeons ago.

For this reason, the study of the dynamics of the mantle, both its evolution and circulation, is critical to our understanding of how the entire planet functions. Processes from plate tectonics and crustal evolution to core freezing and hence the geodynamo are governed, and in many ways driven, by the cooling of the mantle and the attendant phenomenon of mantle convection, wherein hot buoyant material rises and cold heavy material sinks.

7.01.2 A Historical Perspective on Mantle Dynamics

To some extent, the development of the field of mantle dynamics is most closely linked with the history of theories of continental drift and plate tectonics. Although mantle convection was invoked to provide a driving mechanism for continent (or plate) motions, the hypothesis that the mantle flows and circulates predates even that of continental drift (see Schubert et al., 2001; Chapter 1). For example, in 1895, John Perry used the notion of mantle convection to refute the estimate for the age of the Earth given by his former mentor William Thomson, Lord Kelvin (England et al., 2007). However, a great deal of progress on understanding mantle convection also comes from the general study of the physics of thermal convection, not specifically applied to the mantle.

Histories of plate tectonics (or continental drift) are in abundance (e.g., Hallam, 1987; Menard, 1986), and the text on mantle dynamics by Schubert et al. (2001) gives an excellent summary of the history of developments in mantle convection theory in conjunction with plate tectonics. However, it is also important to grasp the historical context and personalities associated with these developments, especially with regard to how the field evolved and to some extent how science in general has been done and is done now. Thus, rather than merely repeat other historical summaries here, I will instead focus on the contributions (pertaining primarily to mantle convection) and professional and personal histories of some of the leading names in the development of the theories of thermal convection and mantle dynamics. Roughly keeping with the structure of this volume, I will concentrate on the origins of the physics, theory, and systematic experiments of convection by visiting Benjamin Thompson (Count Rumford), John William Strutt (Lord Rayleigh), and Henri Claude Bénard. This will be followed by reviewing the lives of some of the pioneers of the quantitative analysis of mantle convection as a driving force of 'continental drift,' namely, Arthur Holmes, Anton Hales, and Chaim Pekeris, and then two leading proponents of convection and its association with the modern theory of seafloor spreading, subduction, and plate tectonics, Harry Hammond Hess and Stanly Keith Runcorn. Apart from hopefully providing an in-depth perspective on the origins of the science of mantle convection, this survey also reveals the intriguing historical ties between many of these famous characters, for example, that Rayleigh had become a professor in the institution that Rumford established, that Holmes had studied under Rayleigh's son, and that Chaim Pekeris was intimately involved with the birth of the state of Israel that Rayleigh's brother-in-law Arthur Balfour helped create.

7.01.2.1 Benjamin Thompson, Count Rumford (1753–1814)

Benjamin Thompson (Figure 1) is one of the more colorful and complex characters in the history of science. He was simultaneously a brilliant observationalist, an egotistical opportunist, and a dedicated social reformer and champion of the poor. His role as spy against the rebelling American colonies on behalf of the British gives him a dubious role in American (although not European) history, in that one of the fathers of thermodynamics also played a role not unlike that of Benedict Arnold.

Rumford is primarily known for his work on the theory of heat as motion – leading eventually to the kinetic theory and thermodynamics – and for working to debunk the caloric theory of heat. Histories of convection will often note that Rumford is credited as possibly being the first to experimentally observe convection; in fact, the study of the mass transport of heat was a significant part of his overall body of work (Brown, 1967), and he wrote an important article on convection in 1797, although the use of the word 'convection' was not coined until much later, by Prout (1834) (see also Ahrens, 1977; Normand et al., 1977).

Benjamin Thompson was born in Woburn Massachusetts in 1753 to a line of Thompsons that can be traced back to a James Thompson who arrived 10 years after the landing of the Mayflower (1620), along with eventual Massachusetts Governor John Winthrop. Thompson's father and grandfather were reasonably wealthy farmers, but his father died young when Benjamin was less than 2 years old. The family farm was inherited by Benjamin's uncle who appears to have treated his nephew well with a significant income, a portion of land, and a high-quality education. As with all the subjects of our histories here, Thompson was a brilliant student, displaying talents in mechanics and natural philosophy; however, he was also known for being a somewhat spoiled child at his family's farm.

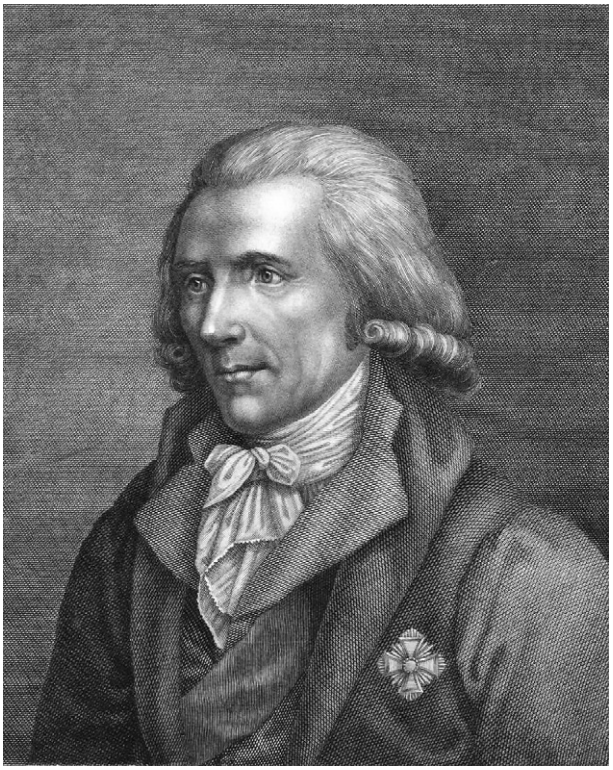


Figure 1 Benjamin Thompson, Count Rumford (1753–1814) (Smithsonian Institution Libraries Digital Collection).

Thompson left school at age 13 for an apprenticeship in retail but continued his studies independently in engineering, mathematics, medicine, and experimental philosophy, along with French, fencing, music, and draftsmanship. He also carried out independent experiments in science, including astronomy, engineering, and anatomy, and nearly electrocuted himself trying to repeat Benjamin Franklin's experiments on thunderstorm electrification.

At age 18, Thompson set out to generate much needed income and turned to tutoring the children of local wealthy families, which led to his being invited by the Reverend Timothy Walker of Concord, New Hampshire, to run a school in his village. Concord was originally known as Rumford, and it is from this town that Thompson was to derive his name upon being ennobled. Thompson courted the Reverend Walker's daughter Sarah, who had recently been widowed following a brief marriage to a much older, wealthy land owner. Within a year, Thompson and Sarah were married (in 1772) when he was 19. Thompson's new wealthy wife facilitated his connections with the British governing class, in

particular by dressing him up in a fine Hussar uniform and parading him about Boston, where he made such an impression on Governor John Wentworth that he was given a British major's commission in the Second New Hampshire Regiment. In 1774, Thompson and his wife had a daughter – also Sarah.

By this time, hostilities between the Colonies and the Crown had been mounting; these included the Stamp Act and subsequent riots (1765), the Boston Massacre (1770), and the Boston Tea Party (1773), which was followed by both the relocation of the Massachusetts capitol from Boston to Salem and the passing of a series of acts called, by the colonists, the Coercive or Intolerable Acts. These events had by now led to the First Continental Congress in Philadelphia in September 1774, which demanded repeal of these acts, as well as calling on civil disobedience and the buildup of local militia called the Minutemen. Eventually, war broke out near Boston in April 1775, at Lexington and Concord.

Benjamin Thompson's acceptance of a British commission in the Second New Hampshire Regiment was thus problematic for several reasons. While British officers resented the presence among their ranks of an inexperienced schoolmaster, the people of Concord and Woburn regarded him as a traitor, and he had to face two trials for 'being unfriendly to Liberty,' in 1774 and 1775, both of which he weathered without formal charges. Nevertheless, Thompson left for Boston in 1775 to offer his services to the British Army, in particular to gather intelligence on rebels by various and nefarious means. In March 1776, the British evacuated Boston and Thompson left with them, abandoning his wife and daughter; he was never to see his wife again, and his daughter not again for 20 years.

In England, Thompson worked for the Secretary of State for the Colonies, and he rose rapidly to other prominent positions in the administration of colonial rule. During this time, around 1778, his studies of force and heat associated with gunpowder explosions and ballistics of large guns led to his election as Fellow of the Royal Society in 1781.

Thompson also briefly went back to the American colonies in 1781 to take up command in the King's American Dragoons in New York. With the end of the war shortly thereafter, Thompson returned to England as a professional soldier and colonel and went to the continent as a soldier of fortune in 1783, landing in Bavaria, eventually to become the advisor to Elector Karl Theodor. As a consequence of his new employment, he was knighted by the British King George III in 1784 to secure his loyalty to England. It was in Bavaria, with the financial and technical backing of his court position, that Thompson did much of his scientific work, especially on the insulating properties of materials and on transport of heat by fluids (i.e., convection). He also did much in terms of military, educational, social, and economic reform in Bavaria, all for which he was rewarded by being made a Count of the Holy Roman Empire in 1792; he assumed the title of Count Rumford adopting the original name of the New Hampshire village where his rapid climb began.

Rumford's frenetic activity and modern innovations were met with great resistance by the established order who eventually held sway with the Elector and caused Rumford to be removed back to London as a minister to the English Court. However, due to a miscommunication with the British Crown, he arrived jobless in London. Unable to remain idle, Rumford established the Royal Institution in London, which employed natural philosophers like Thomas Young and Humphrey Davy to perform experiments and give public lectures. However, his dominating style and large ego led to battles with institute managers, and he eventually left both the Royal Institute and England in 1802. He flitted for political reasons between Bavaria and France, until the invasion of Bavaria by Austria in 1805 sent him permanently to France, where he hoped to gain favor with Napoleon Bonaparte. In France, he married the widow of the 'father of chemistry' Antoine Lavoisier who had been sent to the guillotine at the height of the French Revolution in 1794. However, in France, Rumford fought scientifically with Laplace and Lagrange and separated from Madame Lavoisier de Rumford. He died in August 1814 and was buried in the village of Auteuil.

Rumford was known widely for his philosophic papers on the nature of heat and mounted perhaps the most coherent assault on the caloric theory. The caloric theory is – as with many failed theories – frequently explained with hindsight as an absurdity. However, it is important to understand that it was a reasonably sophisticated physical and mathematical theory that made distinct quantitative predictions about the nature of both heat and matter, which were in fact borne out by experiments. Caloric was considered the medium or 'fluid' that transported and retained heat within matter. But more specifically, caloric was thought to provide the repulsive force – by providing a caloric atmosphere around each atom – within matter to keep it from collapsing under the influence of gravitational attraction. The mathematical predictions of the caloric theory were verified by the experiments of Dulong and Petit on, for example, thermal expansion. Similarly, predictions about the dependence of specific heat on temperature, phase changes (in which caloric is either absorbed or released), adiabatic heating under pressurization (wherein caloric is squeezed out), and heat conduction as driven by caloric potential all confirmed the caloric theory. Indeed, the notion of caloric substance remains a relic in our discussion of heat as 'flowing' or existing as quantities with certain densities such as specific heat or latent heat (Brown, 1967).

Thus, Rumford's assault on caloric theory was not on a ludicrous model that could be easily dispatched by simple thought experiments. However, it did conflict with certain observations, a few key ones Thompson was responsible for. Most notably, these included his work on frictional heating, in particular his observation that during work on boring out cannons, heat is created with no evident source of caloric from other matter and also the propagation of heat in a vacuum by radiation (as caloric was thought to exist only in proximity to atoms).

Rumford's observations of thermal convection. Rumford's work on thermal convection results in a famous paper that was published in multiple venues in 1797 (Brown, 1967; Thompson, Benjamin (Count Rumford), 1797) during his tenure in Bavaria. Much of this work involved the investigation of the insulating properties of matter, which he was researching as part of an effort to improve cold-weather clothing for field soldiers, as well as the poor and underclassed. But this work was also relevant in his attack on caloric theory in that it

showed that heat could be transported by the motion of mass itself, and if such mass flux did not occur, the material was an insulator (or nearly so). This also prompted him to note that the internal motion of fluid particles also provided a ready source of heat, and this was a precursor to his theory of heat as particle motion.

Rumford's most explicit observation of convection itself occurred while examining the 'communication of heat' by taking oversized thermometers filled with liquids (including alcohol derived from wine) and subjecting them to heating. Upon leaving one thermometer in a window to cool, Rumford noticed rapid fluid motion that was well delineated by dust particles (which had been introduced because he had let the tubes sit open for 2 years without cleaning) illuminated by sunlight:

> I observed an appearance which surprised me, and at the same time interested me very much indeed. I saw the whole mass of the liquid in the tube in a most rapid motion, running swiftly in two opposite directions, up and down at the same time. (Thompson, Benjamin (Count Rumford), 1797)

On closer inspection, Rumford noticed a regular circulation of upwelling along the axis of the tube and downwelling along the cooler glass boundary. He further found that dousing the tube with ice water hastened the motion but that the circulation eventually ceased as the entire thermometer reached room temperature.

Rumford reasoned that heat in liquids and gases was only carried by particle motion, and to prove this, he contrived experiments in which the fluid was subjected to thermal gradients but the particle motion was obstructed or 'embarrassed.' He compared heat transport in water when pure with that when mixed with substances such as eider (duck) down and stewed apples, which would 'impair its fluidity.' Rumford concluded that "Heat is propagated in water in *consequence* of internal motions or that it is transported or carried by the particles of that liquid ...." The manner in which heat transport can be inhibited by impaired fluidity led Rumford to speculate on the role of convective heat transport in nature, in particular that impaired fluidity was God's design, for example, observing that

> ... when we advert to the additional increased viscosity of the [tree] sap in winter, and to the almost impenetrable covering for confining Heat which is formed by the bark, we shall no longer be at a loss to account for the preservation of trees during winter. (Thompson, Benjamin (Count Rumford), 1797)

Rumford had in fact noted an important feature of convection in many natural systems, that is, the concept of self-regulation of heat during convective heat transport. This is something considered also in mantle convection, in particular that high temperatures lead to low-viscosity and rapid convective expulsion of heat, while low temperatures lead to greater effective insulation and eventual buildup of heat.

Count Rumford was in every sense an experimentalist and observationalist, but his contributions to the study of heat marked the initiation of the modern theory of classical thermodynamics, kinetic theory, and heat transport by particle motion both microscopically and through thermal convection.

Further readings, including sources for this brief history, are Brown (1967) and Brown (1999).

7.01.2.2 John William Strutt, Lord Rayleigh (1842–1919)

John William Strutt's (**Figure 2**) various contributions to physics are evident in the number of phenomena named for him, including the Rayleigh scattering, Rayleigh–Jeans criterion, Rayleigh waves, and, within the study of convection, the Rayleigh number. His greatest legacy is perhaps his contribution to the theory of waves in solids and fluids; however, he received the Nobel Prize in Physics for the discovery and isolation of atmospheric argon, which was essentially experimental chemistry.

Although Strutt was one of the few noble scientists actually born into nobility (as opposed to, say, Benjamin Thompson who was ennobled in his own life), the Strutts were not from ancient noble stock. The family could be traced back to 1660, where it was known for milling corn with water-driven mills, a business that eventually established the family's financial standing. In 1761, John Strutt (Rayleigh's great grandfather) purchased what eventually became the family estate of Terling Manor and where Rayleigh would eventually build his laboratory in which he did much of his science. John Strutt's oldest surviving son Joseph Holden was a colonel in the West Essex Militia and a member of parliament (like his father); for his public services, he was offered a peerage by George III. However, for not entirely clear reasons, other than an apparent conviction not to accept personal honors, Joseph Strutt declined the peerage and asked that it be bestowed on his

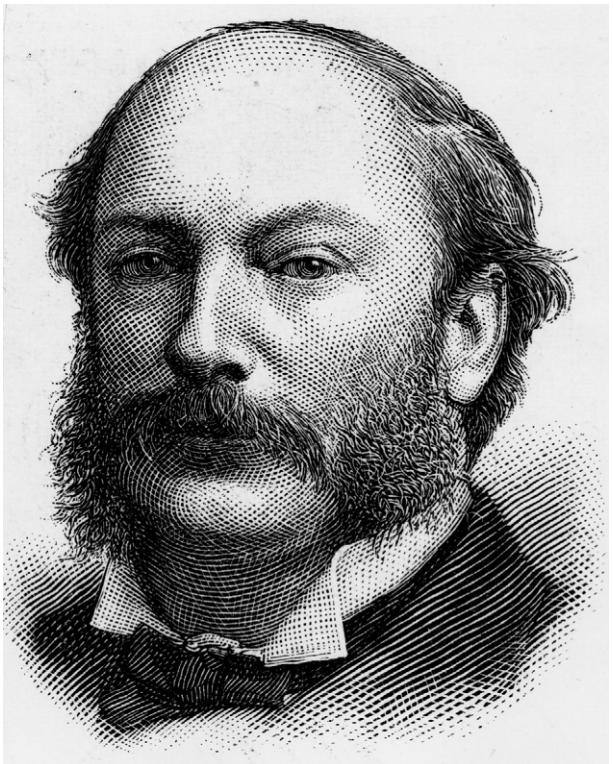


Figure 2 John William Strutt, Third Baron Rayleigh (1842–1919) (AIP Emilio Segre Visual Archives, Physics Today Collection).

wife Lady Charlotte (Fitzgerald) Strutt, whom he had married in 1789. When she died in 1836, the peerage went to Joseph's only son John James, who became the second Baron Rayleigh, during his father's own lifetime. The name Rayleigh was derived from a small market town in Essex (and possibly for no particular reason other than its noble sound). John James married Clara Vicars in 1842 (nearly 30 years his junior); in biographies of Rayleigh, it is usually claimed that his scientific and mathematical talent came not from the Strutts but from the Vicars, who boasted several members of the Royal Engineers and direct descendancy from the brother of physicist Robert Boyle. In November of the year of their marriage, John William Strutt, oldest son and heir to the barony, was born, prematurely. As a child, he did not speak properly till past the age of 3, prompting his grandfather Colonel Strutt to remark that "That child will either be very clever or an idiot" (Strutt, Robert John (the fourth Baron Rayleigh), 1968).

John William suffered poor health throughout his youth and had to be withdrawn from school twice to be tutored privately. He entered Trinity College, Cambridge, in 1861 where he studied rigorous mathematics with, for example, Sir George Stokes, the Lucasian professor of mathematics, and graduated with honors in 1865. After graduating, Strutt forewent the traditional grand tour of the 'continent' and instead visited the United States, which was then in throws of post-Civil War Reconstruction. Upon his return from America, he purchased experimental equipment to set up his own laboratory. There was at the time no formal university physics laboratory in Cambridge. Although much experimental work had already been done by the likes of Michael Faraday, Humphrey Davy, and even Isaac Newton, most of this was outside the university (in particular at the Royal Institution, which, as noted in the preceding section was founded by Count Rumford). It was not until 1871 that James Clerk Maxwell was named the first Cavendish professor of experimental physics and not till 1873 that the Cavendish labs were built.

Strutt's first paper in 1869 was on Maxwell's theory of electromagnetism, for which he received much encouragement from Maxwell himself. In 1871, he married Evelyn Balfour, the sister of Arthur James Balfour. (Arthur James Balfour became the British prime minister from 1902 to 1905 and was the author of the Balfour Declaration (1917) in which Britain formally supported the creation of a Jewish homeland in Palestine during the partitioning of the Ottoman Empire after World War I.)

However, soon after his marriage, Strutt was stricken with rheumatic fever; this prompted an extended tour of Egypt, during which he did much of his work on his famous book *The Theory of Sound*. After his return in 1873, he assumed the title of third Baron Rayleigh on the death of his father and took up residence in Terling where he built his lab in which he did much of his life's scientific work. His book on sound was published in 1877 and 1878 (in two volumes), an achievement that emphasized his lifelong fascination with sound and wave theory in general. During these early years, he also continued to work on sound, electromagnetism, and light, as well as his theory for light scattering and the cause for the sky's blue color.

James Clerk Maxwell's untimely death in 1879, at the age of 48, left the Cavendish experimental chair vacant. The first Cavendish chair had been offered to Sir William Thomson (later Lord Kelvin) in 1871, but he had refused in order to stay in Glasgow, and his decision remained even upon being reoffered the job in 1879. Rayleigh was then offered the position, which he reluctantly accepted because revenue from Terling had declined during the agricultural depression of the late 1870s. Indeed, Rayleigh only held the position until 1884 after which his economic situation had improved and he returned to his residence and work in Terling. His brief tenure as the Cavendish professor was the only academic position he held in his life. However, during his time in Cambridge, he carried out a vigorous program of experimental instruction and research on electrical standards, now having access to lab assistants and more facilities. Also, in his years at Cambridge, he developed a close acquaintance with Sir William Thomson.

After resigning his professorship and returning to Terling, Rayleigh became the Secretary of the Royal Society for the next 11 years (1885–96). In 1887, he became professor of natural philosophy at the Royal Institution (see **Section 7.01.2.1**) at which he gave over a hundred lectures until he left it in 1905.

One of Rayleigh's most prominent discoveries is of argon in the atmosphere. Rayleigh had first suggested the presence of an unknown atmospheric gas because of an apparent anomaly in the density of atmospheric nitrogen relative to that extracted from compounds. He later performed painstaking experiments separating argon from atmospheric gas in 1895. William Ramsay, following up on Rayleigh's earlier suggestion, also conducted experiments to extract argon. Some controversy ensued over whether Ramsay had a right to follow Rayleigh's suggestion and infringe on his research. But even Rayleigh himself agreed that publication of an idea makes it, in effect, public domain, which is very much the prototype of modern publication ethics today. The two shared priority of the discovery for which they each received the Nobel Prize in 1904, Rayleigh's in physics and Ramsay's in chemistry.

Rayleigh became president of the Royal Society from 1905 to 1908 and took on numerous public service roles, chief of which was chancellor of Cambridge University in 1908. He published his complete papers up to 1910 in five volumes; the remaining papers from 1911 to 1919 were published posthumously under the editorship of his son Robert John (who was himself a prominent physicist; see **Section 7.01.2.4**) in 1920. Rayleigh was the author of approximately 450 papers, working essentially up to his death in June 1919; indeed, three of his papers were still in review or in press on the day of his death.

Rayleigh's linear theory of convection. Lord Rayleigh's paper on convection was written very late in his life, in 1916, only 3 years before his death (Strutt, John William (Lord Rayleigh), 1916). He had, however, worked sporadically on the stability of fluid flows since at least the 1880s. Rayleigh's paper on convection was entirely inspired by Bénard's experiments, which had taken place 15 years earlier (**Section 7.01.2.3**). In his 1916 paper, Rayleigh described Bénard's experiments, recognizing two phases, which were an initial transient phase in which an irregular or semiregular pattern is established (with polygons of four to seven sides) and then a second phase of stable and regular polygonal patterns. In a footnote, he also commented that Bénard was perhaps unaware of the work of James Thomson (Lord Kelvin's older brother) who did fundamental work on evaporative convection finding similar polygonal patterns (see Berg et al., 1966). Rayleigh's theory was in

fact focused on the first phase or the transient onset; he did not recognize in the second phase what was later termed 'exchange of stabilities' in that the state to which the perturbations are moving the system is itself stable and nonoscillatory (see Chandrasekhar, 1961). Rayleigh followed a first-order perturbation analysis, although in a rather informal manner. He explored the effect on stability of various individual parameters (e.g., viscosity – considering both inviscid and viscous cases, thermal diffusivity, sign of imposed temperature gradient, gravity, and thermal expansivity), but did not recognize dynamic similarity and dependence on a few dimensionless numbers (see **Chapter 7.04**). Although he did not pose the dimensionless number called the Rayleigh number, it is easily recognizable in his solution for the minimum critical temperature gradient or 'density drop' across the layer (Strutt, John William (Lord Rayleigh) (1916); equations (44) and (46)). Rayleigh also commented on the degeneracy of the linear problem (i.e., mode selection only in terms of wave number squared). But he also recognized that the basic differential operators in the equations allowed wave number pairs that permitted regular polyhedral patterns such as squares and rolls, although he admitted that while hexagons and triangles were obvious, the problem was not immediately tractable (it was, however, later solved analytically; see Chandrasekhar, 1961). Bénard's hexagonal patterns were the cause for some confusion in the comparison of theory and experiment, and they were later inferred to be characteristic of systems where vertical symmetry of the convecting fluid layer is broken (e.g., with temperature-dependent viscosity) (see Manneville, 2006; **Section 7.01.2.3**).

Lord Rayleigh's paper is of course one of the great seminal works in the study of convection, not only for establishing the theoretical framework of the problem and developing the concept of marginal stability but also for inferring pattern selection at convective onset. Further reading and sources for this history can be found in Strutt, Robert John (the fourth Baron Rayleigh) (1968) and Lindsay (1970).

7.01.2.3 Henri Claude Bénard (1874–1939)

While the phenomenon of thermal convection had been observed at least as early as the accounts of Count Rumford (**Section 7.01.2.1**), Henri Bénard (**Figure 3**) is generally credited with having done the first systematic quantitative examination of natural cellular convection. Even given all the problems later found with his experimental results, he is justifiably recognized as the father of experimental studies of convection, which later inspired decades of work not only on convection but also on self-organization and critical phenomena (Manneville, 2006; Wesfried, 2006).

Henri Claude Bénard was born in Lieurey, a small village in Normandy, in October of 1874. His father was a financial investor who died when Bénard was young. After attending schools in Caen and later Paris, Bénard was, in 1894, accepted into the highly competitive Ecole Normale Superieure de Paris, one of the French Grand Ecoles, in a year with about 5% acceptance rate in the sciences. While at the ENS–Paris, he studied with some very notable classmates, for example, physicist Paul Langevin and mathematician Henri Lebesgue, and was witness to the groundswell of outrage over the Dreyfus

Figure 3 Henri Claude Bénard (1874–1939). Reproduced from Wesfried J (2006) Scientific biography of Henri Bénard (1874–1939). In: Mutabazi I, Wesfried J, and Guyon E (eds.) *Dynamics of Spatio-Temporal Cellular Structure – Henri Bénard Centenary Review*, pp. 9–37. New York: Springer, Fig. 2.3, p. 15, with kind permission of Springer Science and Business Media.

Affair. In 1897, he received the *agrégé de physique* (now termed *agrégation*, whereby state teaching credentials are obtained) and began work in the department of experimental physics in the Collége de France, which is unattached to any university and does not grant degrees, but provides open public lectures (similar to the Royal Institution). While there, Bénard was assistant to Eleuthère Mascart and Marcel Brillouin, who were, respectively, grandfather and father of the physicist Leon Brillouin (Wesfried, 2006). While at the Collége de France, he worked on the rotation of polarized light through sugar solutions, which gave him training in the use of optics for measurements of fluid motion; this specific work also led to the second topic of his PhD thesis, which was then required for the French doctorate.

The experiments on convection for which Bénard is well known today (Bénard, 1900, 1901) were the primary part of his PhD thesis, but the topic was arrived at much by accident (see the succeeding text). In the defense of his thesis on 15 March 1901, his committee found the work satisfactory but less than inspiring, criticizing Bénard because, while it was innovative (mostly interesting in its application of optical methods) and above average, it was disappointing in that it provided no general theoretical development to explain the experiments. The report on his defense stated that "though Bénard's main thesis was very peculiar, it did not bring significant elements to our knowledge. The jury considered that the thesis should not be considered as the best of what Bénard could produce" (Wesfried, 2006).

After his thesis in 1900, he briefly settled in Paris, got married (but had no children), and was shortly thereafter appointed to the faculty of sciences in Lyon (1902). In Lyon, Bénard carried out much of his well-known work on vortex shedding around 'bluff' (i.e., prismatic) bodies and further developed his ingenious employment of cinematography in laboratory experiments. In 1910, he was appointed to the faculty of science at the University of Bordeaux in physics under the department head Pierre Duhem.

With the outbreak of war in 1914, and as a former student of the Ecole Normale, which carried with it state obligations, Bénard entered the military and was made an officer. He was put on a military scientific commission wherein he worked on problems relating to food refrigeration under transport and, later, on the military use of optics (e.g., using polarized light for tracking ships and submarines and for improving periscopes).

After the war, Bénard returned to science, and in 1922, he moved to Paris as a professor in the Sorbonne University. In 1928, he became the president of the French Physical Society. In 1929, he participated in the development of the Institute of Fluid Mechanics and in 1930 became professor of experimental physics. For the next decade, he continued to work with various students on convection and vortex shedding. He died in 29 March 1939, slightly shy of 65.

Bénard's experiments on convection. Henri Bénard's first observations of cellular convective motion came in 1898, at the Collége de France, while trying to make a coherer of solid dielectrics. (A coherer is a loose often granular agglomeration of conductors or semiconductors whose conductivity is affected by impingement of radio waves.) During his preparations, he noticed a polygonal pattern in melted paraffin-containing graphite dust. From there, Bénard detoured into a painstaking and systematic study of convection, in particular the difficult task of finding the onset of convection as near to conductive stability as possible. To carry out this task meant eliminating minute thermal fluctuations imposed at the boundaries, and so, Bénard constructed an apparatus composed of a metal container with steam circulation in its walls to provide a nearly uniform isothermal bottom boundary; the top however was exposed to air. Bénard observed different patterns of convection cells, which had polygonal structure of four to seven sides (see also **Section 7.01.2.2** and **Chapter 7.03**) but were predominantly hexagonal. He termed these cells *tourbillons cellulaires* or cellular vortices, although we now refer to them as Bénard cells or convective cells. Bénard was able to measure closed streamlines of particle flow and observed an initial transient state of polygon formation, settling down to a stable hexagonal configuration after some time, which allowed him to measure cell sizes accurately. He also observed convective rolls (or *tourbillons en bandes*, what Rayleigh later referred to as striped vortices), which occurred at low heat flux, and high-heat flux turbulent 'vortex worms.' What is highly notable is the large number of innovative optical techniques Bénard used and developed. He introduced the use of not only cinematography but also particle trajectories, interference fringes due to light reflected off hills and valleys on the convectively warped free surface, and also transmission across the fluid layer, which is essentially the same as the shadowgraph technique commonly used today (see **Chapter 7.03**). The combination of these optical effects in fact allowed him to estimate isotherms with (quite impressively) 0.1 °C contours.

Years later, in 1916, Rayleigh analyzed Bénard's experiments (see **Section 7.01.2.2**) but assumed free-slip top and bottom boundaries (which was an analytically tractable configuration of boundary conditions). Rayleigh's work confirmed some of the patterns observed by Bénard, but not the critical conditions for convective onset (what we now call the critical Rayleigh number) due to the inappropriate boundary conditions. Because of World War I, Bénard was not aware of Rayleigh's work until the 1920s, well after Rayleigh's death. In the late 1920s and early 1930s, Bénard compared his experiments to the theoretical work of Sir Harold Jeffreys who had repeated Rayleigh's stability analysis but with the appropriate mixed boundary conditions of a free-slip top and a no-slip bottom (see Wesfried, 2006). Bénard found that the patterns predicted in Jeffreys work matched some of his experiments, but the conditions for convective onset still did not agree, implying a significant disparity between theory and Bénard's experiments.

Bénard was aware at the time of his first experiments of problems inherent with an open surface, but he was mainly concerned with the fact that some working fluids, especially volatile ones such as alcohol, experienced evaporative convection, which is often coincident with Marangoni (surface tension-driven) convection (see Berg et al., 1966). Bénard thus used fats or oils (i.e., spermaceti) and wax that had higher melting temperature and thus lower vapor pressures. Block (1956), however, repeating Bénard's experiments, later suggested that thermally induced surface tension gradients, rather than thermal buoyancy, were the cause for the observed motions and surface deflection; surface tension-driven convection, now known as Marangoni–Bénard convection, was formally developed by Pearson (1958).

Throughout his life, Bénard continued to draw analogies between the cells of his experiments and natural ones, in some cases incorrectly (e.g., Taylor–Couette rolls). However, he correctly advocated the cellular cause of solar granulation. Moreover, he promoted the idea that cloud streets were due to longitudinal convective rolls aligned parallel to wind; he directed experiments on convection in a tilted layer and ones with a moving top boundary, to find alignment of convection rolls (see Wesfried, 2006, and references therein), essentially identical to what it is referred to in mantle dynamics as Richter rolls (see **Chapter 7.07**).

Additional reading and historical sources for this brief history are Wesfried (2006) and Manneville (2006).

7.01.2.4 Arthur Holmes (1890–1965)

The prospect of convective heat transport in the mantle was alluded to possibly as early as the mid-nineteenth century (e.g., England et al., 2007; Schubert et al., 2001). However, Arthur Holmes (**Figure 4**) can rightly be considered one of the founders of the physical theory of mantle convection as it pertains to the driving mechanism of continental drift. Even so, Holmes is still perhaps best known for championing the science of radiometric dating to infer the age of the Earth and for establishing the geologic time scale (Lewis, 2000).

Arthur Holmes was born in Gateshead, in the northeast of England, in 1890. As a precocious teenage student, he was strongly influenced by a teacher who introduced him to both physics and geology through the writings of William Thomson (Lord Kelvin) and the Austrian geologist Eduard Suess. At the age of 17, he went to London to study physics at the Royal College of Science, which was then being absorbed into the new Imperial College London, but then changed directions to geology. However, his background and interest in physics would serve his geologic ventures throughout his life. In his final year at the Imperial College London, he studied the novel

Figure 4 Arthur Holmes (1890–1965). Used with permission from the Arthur Holmes Isotope Geology Laboratory at Durham University.

and exciting phenomenon of radioactivity and the prospect of radioactive dating of rocks, under the new young professor Robert John Strutt, the son of Lord Rayleigh. Strutt had himself been involved in a vigorous and public debate over the age of the Earth with his father's old friend, Lord Kelvin (Lewis, 2000).

The debate over the age of the Earth is traditional historical fare in geoscience (see Lewis, 2000). Suffice it to say that the controversy swirled about both Bishop Ussher's biblically inferred age of 6000 years and Lord Kelvin's cooling age of 20 My, neither of which could be reconciled with geologists' observations of sedimentation rates that required the oldest rocks to be no less than 100 My old. However, the discovery of radioactivity in rocks and the resolution of radioactive half-lives suggested a nearly direct measure of rock ages. (It also provided a heat source for keeping the Earth from having to cool from a recent molten state.) Several scientists attempted to develop the technique of radiometric dating, including Strutt and Ernest Rutherford. But, in the end, Yale scientist Bertram Boltwood identified lead as the final product of the uranium–radium decay series; since lead was a nonvolatile and thus a nonleaking daughter product – unlike helium – its concentration could be reliably measured. Boltwood used this decay series to infer the age of rocks in both Connecticut and Ceylon (now Sri Lanka) and determined ages of these rocks between 535 My and 2.2 Gy and published these results in 1907 (see Turekian and Narendra, 2002). Using Boltwood's method, Holmes similarly dated Devonian rocks and arrived at an age of 370 My; his results were presented (in his absence) at the Royal Society in 1911.

Indeed, by 1911, Holmes was already in Mozambique where he had taken a job in mineral prospecting because his scholarly stipend in London was not enough to live on. However, he was there only 6 months, having fallen seriously ill with malaria. He returned to London to take a job as demonstrator in the Imperial College London where he continued to develop geochronological methods. In 1913, Holmes published his first book, at the age of 23, *The Age of the Earth* where he calculated the planet's age to be 1.6 Gy old, which was less than Boltwood's estimate. Although geologists were, on the whole, relieved that the 6000 years and 20 My ages were proven wrong, many still held fast to the 100 My date and were reluctant to accept the radiometrically inferred ages, probably because many did not understand the new physical principles of radioactivity (Lewis, 2000, 2002).

Holmes married Margaret (Maggie) Howe in 1914 and he continued working as a demonstrator in the Imperial College London through World War I. After the birth of his son Norman, in 1918, Holmes found that he was unable to support his small family on his demonstrator's salary and thus took a job prospecting for oil with a company in Burma in 1920. However, the company soon went bankrupt. Moreover, Holmes' 4-year-old son Norman fell ill with dysentery and, despite available medical attention, died in 1922. Holmes and his wife soon returned to England, impecunious and grieving. His misfortunes continued, however, and he could not find a position until 1924 (in the meantime running a curio shop for income), when he was offered a professorship at the University of Durham to build a new geology department.

Holmes remained at Durham for nearly 20 years. In that time, he produced his now famous papers on mantle convection as the cause for continental drift (see the succeeding text) and wrote his text *Principles of Physical Geology*, the first edition of which appeared in 1944. However, also during those years, Holmes' wife Maggie died (in 1938), but he was soon thereafter remarried to Doris Reynolds, a fellow geologist.

Holmes left Durham in 1943 to assume the Regius Professorship at the University of Edinburgh where he remained until retirement in 1956. While at Edinburgh, he continued to work on refining the age of the Earth and developing the geologic time scale, in addition to other geologic pursuits. He received various high honors (e.g., the Wollaston and Penrose Medals in 1956 and the Vetlesen Prize in 1964), primarily for his work on the geologic time scale, not on continental drift and convection. In the early 1960s, evidence for plate tectonics was mounting, especially with the Vine–Matthews work in 1963 on seafloor spreading. Holmes' sense of vindication is evident in his revised text, which was published the year he died, in 1965.

Holmes and mantle convection. By the late 1920s and early 1930s, the debate over the age of the Earth had given way to a new controversy over the theory of continental drift. As is well known, the German meteorologist Alfred Wegner had proposed his idea based largely on geographic evidence (Hallam, 1987; Wegener, 1924). But he also proposed that continents plowed through oceanic crust like ships, and the driving force was due to centrifugal effects. Although Sir Harold Jeffreys had done some of the most fundamental work on convection

theory (see Schubert et al., 2001; **Section 7.01.2.3**), he had argued that there were no available forces sufficient to deform the Earth's crust during continental drift. Holmes, on the other hand, supported the idea of continental drift but, along with Bull (1931), proposed that subsolidus convection in the mantle – powered by heat production from radioactive decay – was instead the driving mechanism for continental breakup, seafloor formation (not spreading), crustal accumulation at convergence zones, and continental drift (Holmes, 1931, 1933). Holmes' ideas of subsolidus convection were similar to present-day understanding. Some notable differences his model has with contemporary mantle convection theory are that Holmes believed that mantle flow would establish jets and prevailing winds as in the atmosphere, although these in fact arise through the combination of convection and planetary rotation; as discussed in **Chapter 7.02**, the effects of rotation are not significant in mantle circulation. Moreover, Holmes also proposed that seafloor formation was associated with deep and active mantle upwellings, which was later proved to be unlikely even as early as the 1960s (see **Section 7.01.2.7**). Holmes' theories of convection were, like Wegener's theory of continental drift, rebuffed and ignored, although Holmes continued to teach these ideas while at Durham and Edinburgh; indeed, his famous text contains a final chapter discussing his view of continental drift and convection.

Further reading and sources for this section can be found in Hallam (1987), Lewis (2000, 2002), and Schubert et al. (2001).

7.01.2.5 Anton Hales and Chaim Pekeris

Arthur Holmes is largely seen as a visionary in being the champion of convection as the driving mechanism for continental drift. However, near the same time as his first papers, two important papers on mantle convection were published, that is, by Anton Hales (Hales, 1936) and Chaim Pekeris (Pekeris, 1935). Both of these used modern fluid dynamic theory not only to estimate the conditions for convection, as Holmes had done, but also to calculate the finite-amplitude velocity and stresses of convective currents and to compare them with predictions from gravity observations. Not only were their theories fluid dynamically sophisticated, but also their predictions were born out 30 years later in measurements of plate motions. Moreover, their respective papers were precursors to the modern analysis of how convection is reflected in gravity, geoid, and topography. Thus, these two authors warrant some discussion. Both Hales and Pekeris were perhaps better known for their lifelong contributions outside of mantle dynamics (e.g., seismology), but they both played important roles in the growth of geophysics in the twentieth century.

7.01.2.5.1 Anton Linder Hales (1911–2006)

Anton Hales (**Figure 5**) was born in Mossel Bay, in the Cape Province of South Africa, in March 1911. As with all our historical subjects, Hales showed an early aptitude and talent for science and graduated from the University of Cape Town with a BS in physics and mathematics, at the age of 18, and an MS at 19. He then, in 1931 at the age of 20, took up a post as a junior lecturer in mathematics at the University of Witwatersand in Johannesberg (Lilley, 2006). However, after only one more year, he received a scholarship to study at Cambridge. Although intending to study quantum mechanics, he was convinced by Sir Basil Schonland (a senior lecturer in physics at the University of Cape Town) that this was the 'wrong choice' (Lambeck, 2002) and that he should instead study geophysics. Thus, while in Cambridge, Hales studied with Sir Harold Jeffreys and interacted with Keith Bullen and received his BA in mathematics from St. John's College in Cambridge in 1933.

Hales returned to South Africa where he resumed his post as junior lecturer and eventually senior lecturer in applied mathematics at the University of Witwatersand. While there, he also carried out research primarily in seismology for which he received his PhD from the University of Cape Town in 1936. That year, Hales married Marjorie Carter with whom he had two sons, James and Peter (Lilley, 2006).

At the outbreak of the World War II, Hales' scientific career was put on hold and he served as an engineering officer in the North African campaign. After the war, Hales left his lecturer position to become a senior researcher at the Bernard Price Institute of Geophysical Research at the University of Witwatersand where he worked on the development of seismic and gravity measurement methods. However, in 1949, he left BPI for a professorship in applied mathematics and the head of the Mathematics Department at the University of Cape Town. During these years, he went briefly to Cambridge (1952) to receive a masters degree. In 1954, he returned to Witwatersand as director of BPI and professor of geophysics. While being the director of BPI, he continued to push the development of geophysical methods, including paleomagnetism, and he was involved with some of the first measurements anticipating plate tectonic motions by looking at pole paths in South Africa,

Figure 5 Anton L. Hales (1911–2006). Used with permission from the Department of Geosciences of the University of Texas at Dallas.

which began to convince Hales of the validity, after all, of continental drift (Lambeck, 2002). It was also during his time at BPI, in 1957, that Marjorie, his wife of 21 years, died.

In 1962, Hales left South Africa for the United States to become the founding director of the Geoscience Division for the Southwest Center for Advanced Studies, later to become the University of Texas. That year, he also remarried Denise Adcock with whom he had two more sons, Mark and Colin. While in Texas, Hales continued to build a powerful research institute and made major contributions to seismic studies of the crust and upper mantle.

In 1973, at the age of 62, Hales was convinced (most notably by Ted Ringwood and John Jaeger) (Lambeck, 2002) to move to the Australian National University to become the founding director of the Research School of Earth Sciences (RSES) where he served until 1978. Hales was an active and unique director in that he minimized departmental structure and bureaucracy and pushed his scientific staff to work globally rather than on regional Australian studies, thereby establishing the school's reputation as one of the world's foremost Earth science institutes. Moreover, under his directorship, the SHRIMP ion microprobe – then a very new and expensive technological advance in geochemical analysis – was developed.

Hales retired from ANU in 1978 and returned to the University of Texas as professor of geophysics. He retired from the University of Texas shortly thereafter in 1982 and returned to ANU and RSES to resume his position as emeritus professor until 2002. By the time of his retirement, he had been made a fellow of the Royal Society of South Africa, the American Geophysical Union, and the Australian Academy of Sciences; in 2003, he was given the Centenary Medal from the Australian government. Hales stayed in and around Canberra for the remainder of his life. However, in 2004, his son Mark from his second marriage was tragically killed in a car accident (Lilley, 2006), only 2 years before Anton Hales himself passed away, in December of 2006, at the age of 95.

Hales *and convection in the mantle*. Anton Hales' work on the viability of mantle convection occurred during his doctoral studies but was unrelated to his dissertation research; his paper on the subject was published in 1935 before receiving his PhD. The problem he examined, suggested to him by Harold Jeffreys, concerned the plausibility of convection with regard to whether the buoyant stresses driven by a mantle of a certain viscosity and heat input were consistent with those inferred by the gravity anomalies measured during the famous submarine gravity surveys of Vening Meinesz. To calculate convective stresses, Hales estimated convective velocities by equating mid-mantle advective heat transport with the conductive surface heat transport necessary to remove the net radiogenic heat production. Hales' relationship for his velocities could be shown to yield very plausible tectonic velocities, as mentioned by Jeffreys himself in the later editions of his famous text (Jeffreys, 1959). (Similarly, as shown in Bercovici (2003), a simple balance of the net advective heat extraction by slabs of a given mean temperature anomaly against the known mantle cooling rate predicts quite readily slab velocities of 10 cm year^{-1}. See also **Chapter 7.07**.) Hales used these velocities to estimate the stresses of convective currents and compared these with the stresses necessary to support mass anomalies inferred from gravity measurements. Although Jeffreys no doubt believed that convective stresses in a stiff mantle would be far in excess of those predicted by gravity, Hales showed that the stress estimates were in fact very close and easily permitted a convecting mantle. Jeffreys 'communicated' (i.e., sponsored) his former student's findings to the Royal Astronomical Society for publication (Hales, 1936).

Further readings, including sources for this brief history, are Lambeck (2002) and Lilley (2006).

7.01.2.5.2 Chaim Leib Pekeris (1908–93)

Chaim Leib Pekeris (**Figure 6**) was born in Lithuania, in June 1908, in the town of Alytus, where his father was a baker. He was the oldest of five siblings and had two brothers and two sisters. As a youth, he was (unsurprisingly) precocious in mathematics and was apparently teaching high-school math by age 16. In the 1920s, he and his two brothers emigrated to the United States with the help of family and friends already in America (Gilbert, 2004; Gillis, 1995). The three Pekeris brothers became American citizens and continued their education in the United States. In contrast, one of his sisters moved to Palestine in 1935 as a Zionist. The remaining sister and their parents were, however, later murdered by anti-Semites in Alytus during the Holocaust.

Chaim Pekeris entered the Massachusetts Institute of Technology (MIT) in 1925 to study meteorology and obtained his BS in 1929. He stayed on for graduate work with Carl-Gustaf

Figure 6 Chaim L. Pekeris (1908–93). Reprinted from Gilbert F (2004) Chaim Leib Pekeris, 15 June 1908–24 February 1993. *Biographical Memoirs* 85: 217–231, with permission from the National Academies Press Copyright 2004, National Academy of Sciences.

Rossby and obtained his doctorate in 1933. During his graduate career, he was also a Guggenheim fellow and studied meteorology in Oslo. After finishing graduate studies, he became an assistant geophysicist in the Department of Geology at MIT and from there rapidly transitioned away from meteorology. He also received a Rockefeller Foundation Fellowship in 1934 and married Leah Kaplan the same year.

Pekeris had been hired at MIT by Louis Slichter who had himself been hired by MIT to establish a geophysics program. Pekeris was his first hire and the second was Norman Haskell, a new PhD from Harvard. The combination of Pekeris and Haskell during the early 1930s made important contributions to the burgeoning problem of continental drift and mantle flow. Indeed, Haskell performed the first analysis of Fennoscandian uplift from which the viscosity of the mantle was initially estimated (Haskell, 1937). In conjunction, Pekeris did a 'hydrodynamic' analysis of thermal convection in the Earth's mantle, which led to an accurate first-order estimate of mantle flow and velocities (see the succeeding text).

Chaim Pekeris also worked on free oscillations, but in particular of stellar atmospheres, although this laid the foundation for his study of free oscillations of the Earth for which he was perhaps best known (Gilbert, 2004). He further studied pulse propagation and inverse problems in sonar sounding and was promoted to associate geophysicist in 1936. From 1941 to 1945, Pekeris worked for the Division of War Research at the Hudson Laboratories of Columbia University, again studying the propagation of acoustic pulses and waves; he continued there as director of the Mathematical Physics Group (1945–50) and had a joint appointment at the Institute for Advanced Study at Princeton. For his war research, he was given the title of honorary admiral (Gilbert, 2004).

After the war, Pekeris continued his scientific research in several areas, including microwaves, atomic physics, and explosive sound propagation through a fluid–fluid interface, which led to a seminal paper on normal modes and dispersion. He also produced the first theoretical derivation for the critical Reynolds number for onset of instability in pipe flow (Gilbert, 2004).

In addition to his scientific work, Pekeris was involved with assisting in the 'birth' and stability of the new state of Israel by aiding in the transfer of US military surplus to Palestine (Gilbert, 2004). In 1950, Chaim Weizmann, the first president of Israel, convinced Pekeris to move to the Weizmann Institute of Science in Rehovot, Israel, to be the founding chair of the department of applied mathematics.

At that time, conditions for the university in Rehovot were extreme since Israel was in an almost constant state of war, with many students and faculty serving in the military. The situation likely contributed to Pekeris' pragmatic approach of pursuing the development of applied math and physics through a computational effort (Gillis, 1995). Thus, part of his negotiation to come to the Weizmann Institute of Science was to be given funds to build one of the world's first digital computers for scientific studies, the WEIZAC (Weizmann Automatic Computer), which was completed in 1955 and whose design was based on the von Neumann machine. The WEIZAC's first use was to solve Laplace's equations for the Earth's ocean tides for realistic continental boundaries; this was a major accomplishment showing the power of computers to turn theory into 'modeling' (Gillis, 1995).

Pekeris continued to build and recruit for the applied mathematics department and mentored many grad students who went on in science and other faculty positions. He was also involved in establishing the first Israeli geophysical survey that led to the discovery of oil in Israeli territory. In 1952, he was elected to the US National Academy of Sciences. While in Israel, he continued to work on atomic physics (on the ground states of helium) and wave propagation and free oscillations in both stars and the Earth, and he was able to test his free oscillations theory with data from the giant 1960 Chile earthquake. He and his students continued the development of computing synthetic seismograms from generalized ray theory, which was computationally intensive and required a computing upgrade from WEIZAC to the more powerful GOLEM series of Israeli-developed 'super computers.' He continued to work on all these problems of atomic physics, seismology, tides, free oscillations, and hydrodynamics until his retirement at the age of 65 in 1973, which was also the year that his wife Leah passed away. Even after retirement, he continued to do research, for example, publishing again on the physics of ocean tides in 1978. Near and during this postretirement time, he was recognized for his lifelong contributions and was elected to various societies and given prizes such as the Vetlesen Prize (1973), the Gold Medal from the Royal Astronomical Society (1980), and the Israel Prize (1981). In 1990, the Mathematical Geophysics Meeting was held in Jerusalem in honor of Pekeris and his contributions. In February of 1993, Chaim Pekeris died as a result of injuries from a fall in his home in Rehovot. The following year, the Weizmann Institute of Science began the annual Pekeris Memorial Lecture.

Pekeris' model of mantle convection. Chaim Pekeris' paper on the viability of mantle convection (Pekeris, 1935) was in many ways one of the first truly sophisticated analyses of mantle convection. His basic hypothesis was to examine the convective circulation caused by lateral thermal gradients associated with the difference between a warm subcontinental mantle and a cooler suboceanic mantle, both estimated from crustal thickness and mantle heat production. Pekeris also used Haskell's value for mantle viscosity and calculated that the convective velocities were of the order of 1 cm year^{-1}, which is a perfectly plausible tectonic velocity. But Pekeris' theory also laid the groundwork for other effects, including accounting for convection in a deep spherical shell, the effect of convection on distortion of the surface, and the net effect of convective density anomalies and surface deflections on the gravity field, thereby predating modern analysis of geoid and topography by 50 years. Pekeris' paper was published 5 months before Anton Hale's paper, although they were both received at the same time (December 1935) and were similarly communicated by Harold Jeffreys to the Royal Astronomical Society. As with Hales, Pekeris was not to examine mantle dynamics again since it is likely that this field was perceived as too speculative and without promise.

Further readings, including sources for this brief history, are Gilbert (2004) and Gillis (1995).

7.01.2.6 Harry Hammond Hess (1906–69)

Harry Hess (**Figure 7**) was one of the giants of the plate tectonics revolution, not only through his legacy of careful

sea-going observations but also in his landmark paper (Hess, 1962) hypothesizing the essence of seafloor spreading and subduction. Hess was also (along with Keith Runcorn; see Section 7.01.2.7) one of the leading proponents, at the dawn of the plate tectonics revolution, for the mantle convection-driving mechanism of plate motion.

Harry Hammond Hess was born in New York City in May 1906. At the age of 17, he entered Yale University to study electrical engineering but switched to geology and, despite purported failures at mineralogy, graduated with a BS in 1927.

After graduating from Yale, he worked as a mineral prospector in Northern Rhodesia for 2 years. He then returned to the United States where he started graduate school at Princeton. For his PhD, he worked on ultramafic peridotite, thought to be part of the mantle. During graduate school, he also took part in some of Felix Andries Vening Meinesz's submarine gravity surveys, particularly of the West Indies island arc, which later inspired him for further marine surveys while in the Navy. Hess finished his graduate studies and received his PhD in 1932. After graduation and brief appointments at Rutgers and the Geophysical Laboratory at Carnegie, he returned to Princeton where he joined the faculty in 1934 and was to remain there for the rest of his life, other than brief visiting professorships at the University of Cape Town from 1949 to 1950 (where he likely interacted with Anton Hales) and Cambridge in 1965.

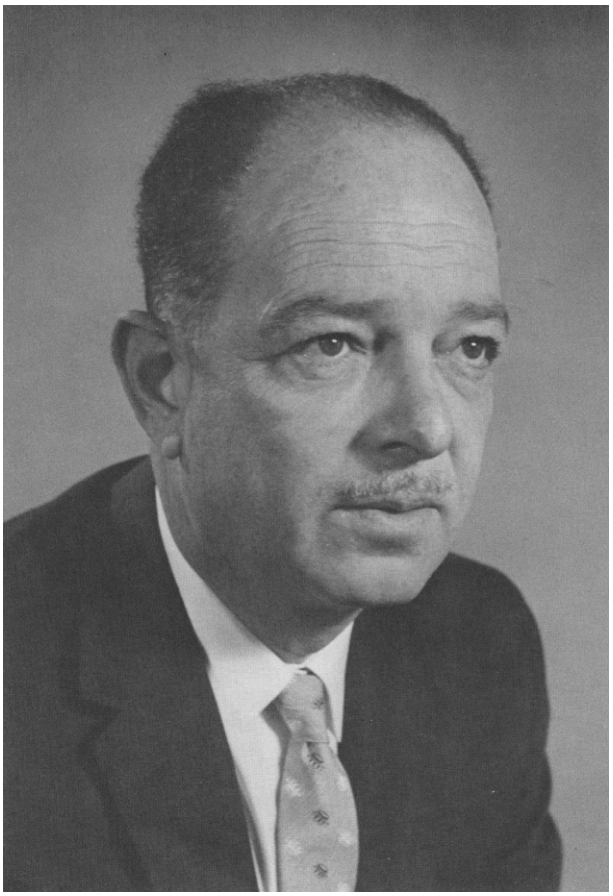


Figure 7 Harry H. Hess (1906–69). Reproduced from Shagham et al. (eds.) (1973) *Studies in Earth and Space Sciences: A Memoir in Honor of Harry Hammond Hess*. Geological Society of America Memoir 132. Boulder, CO: Geological Society of America.

To continue the submarine activities he started with Vening Meinesz, Hess arranged in the 1930s a commission as a lieutenant in the US Navy reserve. After the attack on Pearl Harbor in December of 1941, he was called to active duty. He was first involved with enemy submarine detection in the North Atlantic and developed a technique for locating German submarines. He volunteered for hazardous duty to complete the submarine detection program by joining the submarine decoy vessel USS Big Horn. He was later made commanding officer of the USS Cape Johnson that was a transport ship. Hess was involved with some of the major island-hopping landings of the Pacific theater, such as the landings at the Marianas, Leyte, Linguayan, and Iwo Jima. However, Hess also continuously ran the sonar echo sounder to detect bathymetry en route to these landings, thereby collecting seafloor profiles all across the North Pacific. These surreptitious surveys led to the discovery of 'guyots,' flat-topped seamounts, which Hess named after Swiss geographer Arnold Guyot who founded the Princeton department. Hess inferred these guyots to be islands that had been eroded to sea level but fallen below sea level due to seafloor migration, hence providing one of the first major clues of seafloor mobility. Such seafloor mapping continued under the newly formed Office of Naval Research, and this led to the discovery of the mid-ocean ridge system, which was also found to have, in many locations, rift-shaped valleys running along its length.

After the war, Hess returned to Princeton and initiated and directed the massive multinational Princeton Caribbean Research Project, which set out to perform a comprehensive exploration of Caribbean geology, producing more than 30 PhDs in the process. Hess was made department chair in 1950, a position he kept for 16 years, and during which he led a large expansion of the Princeton department. In 1952, he was elected to the National Academy of Sciences (the same year as Chaim Pekeris' election) and also served on various national advising committees. In the late 1950s, he and Walter Munk initiated the Mohole Project to drill through the ocean crust into the mantle; Hess continued his work on Mohole until 1966, leading to technical breakthroughs that paved the way for the Deep Sea Drilling Project.

Harry Hess continued to serve his department and as national adviser through the 1960s. He was involved with the development of the national space program and was on a special panel appointed to analyze rock samples brought back from the Moon by Apollo 11. He received numerous honors and awards: apart from National Academy of Sciences membership, he also was given the Penrose Medal in 1966 and elected to the American Academy of Arts and Sciences in 1968, and in 1969, just months before his death, he was given an honorary doctorate by Yale.

In late August, 1969, Hess was in Woods Hole Massachusetts, chairing a meeting of the Space Science Board of the National Academy, which was discussing the scientific objectives of lunar exploration only a month after the successful Apollo 11 mission. During the meeting, on August 25, Hess suffered a fatal heart attack. He was buried in Arlington National Cemetery. He was posthumously given the NASA Distinguished Public Service Medal, and the AGU established,

as one of its primary awards, the Harry H. Hess Medal in his honor (Dunn, 1984; Leitch, 1978).

Hess, plate tectonics, and mantle convection. In 1960, Harry Hess wrote an internal report to the Office of Naval Research propounding his hypothesis of seafloor spreading. Years of experience in seafloor surveying led him to believe that mid-ocean ridges had rift valleys; his accumulated knowledge of sedimentology and petrology also convinced him that neither sediments nor fossils on the seafloor were ever more than a few hundred million years old. Hess was essentially convinced by Wegener's observations of continental breakup, and he proposed the idea that the ocean crust diverged away from linear ridges of volcanic activity, later known as seafloor spreading, and that sediments were swept into trenches. He also proposed that ocean crust was subducted, but continental crust tended to scrape off sea sediments to make mountain ranges. In 1962, Hess republished the same paper in a peer-reviewed volume (Hess, 1962), and this paper became one of the most famous and highly cited papers in geoscience and was a landmark in the plate tectonics revolution. However, evidence and verification of the idea did not come until the Vine–Matthews study of magnetic seafloor lineations in 1963. The idea that the seafloor traveled like a conveyor belt essentially satisfied the main objections to continental drift.

At the same time that Hess proposed seafloor spreading, he also concluded, in the same paper, that the driving mechanism for the surface motion was mantle convection, based on observations of gravity anomalies and studies of peridotites, which being ultramafic were assumed to upwell from the mantle. His convection postulate was not based on fluid dynamic analysis but on a synthesis of observations and physical intuition about the ramifications of seafloor spreading and seafloor destruction at trenches. His intuition about the importance of subduction as convection was largely correct, although he, like Holmes, assumed that ridges involved active upwelling (which he also thought carried much more water than is presently known to occur) and that ridges were hence pried apart by convection (Hess, 1962).

Additional reading and historical sources for this brief history are Buddington (1973), Leitch (1978), and Dunn (1984).

7.01.2.7 Stanley Keith Runcorn (1922–95)

Keith Runcorn (**Figure 8**) is a unique figure in that he was at the forefront of major discoveries in paleomagnetism that led to the modern theory of plate tectonics and was also an early and vigorous proponent for the theory of mantle convection, as a driving mechanism not only for continental drift but also in other planets, particularly the Moon.

Stanley Keith Runcorn was born in Lancashire, England, in November 1922 and was educated there as a youth. While an intellectually active young man, he was more interested in history and geography, although he was eventually persuaded by his father and school headmaster to pursue science, particularly astronomy (Collinson, 1998). At age 18, in 1940, he went to Cambridge to study electrical engineering and graduated in 1943. He subsequently joined a telecommunications research firm in Worcestershire to work on radar during the remainder of World War II.

Figure 8 S. Keith Runcorn (1922–95). American Geophysical Union, Courtesy, AIP Emilio Segre Visual Archives.

After the war, in 1946, Runcorn took an assistant lectureship in physics at Manchester University to initially work on cosmic radiation but then moved to study stellar and planetary magnetic fields. The first dynamo theories by Elsasser and Bullard had just been proposed, and Runcorn's first work was to test the core origin of the geomagnetic field by examining the variation of horizontal field strength in coal mines, which he found to increase with depth, thereby lending support to the core origin of the field; this work led to Runcorn's PhD degree in 1949 (see Collinson, 1998).

In 1950, Runcorn moved to the Cambridge Department of Geodesy and Geophysics and worked on remanent magnetism of rocks of different ages to infer polar wander paths for both Great Britain and North America, which were found to differ. After testing and eliminating the possibility of large uncertainty in the data, Runcorn became convinced that the variation in paths was due to continental drift, and this set of observations in the end formed one of the cornerstones for the advent of plate tectonics (see Girdler, 1998).

In 1956, Runcorn left Cambridge for the chair of physics at King's College, University of Durham at Newcastle upon Tyne. He was to remain there until his retirement from the British system in 1988. He quickly established a geophysics program within the Department of Physics (largely by moving Cambridge colleagues with him) and continued to work on and foster paleomagnetic studies, with considerable field work of his own in the Western United States. In this period, he also

established his well-known reputation for extensive travel and departmental absence, earning the facetious title of 'Theoretical Professor of Physics' (Collinson, 1998).

It was during his time in Newcastle that Runcorn began his work on mantle convection (see in the succeeding text), which also led to his growing core hypothesis that had implications for the change of length of day. These ideas prompted Runcorn to move toward the study of coral growth rings, which demonstrated that he had no fear of moving into a different field, even if it was biology. He also continued work through the 1960s on ocean currents and tides but then eventually landed on the study of the Moon, which, given the advent of space exploration and the Apollo program, was a field he would pursue for the remainder of his life. Runcorn continued to work on ideas of convection and geomagnetism in the Moon and on various other topics including Chandler wobble, Jupiter's rotation, and Mars geodesy. In the late 1980s, before his retirement, he also became interested in variations in the gravitational constant G, but this pursuit was eventually found to be fruitless. Runcorn retired from Newcastle in 1988 and took a part-time chair at the University of Alaska at Fairbanks, while also keeping a position in physics at the Imperial College London. During his life, Runcorn had been elected a fellow of the Royal Society (1965) and received the Gold Medal of the Royal Astronomical Society (1984), the John Adam Fleming Medal of the AGU (1983), the Vetlesen Prize (1971), the Alfred Wegener Medal (1987), and several other honors.

In 1995, on the way to the AGU meeting in San Francisco, Keith Runcorn stopped off in San Diego to give a seminar and discuss Galileo Orbiter results with a colleague. On December 5, he was found dead, the victim of a violent homicide, the perpetrator of which was later apprehended, found guilty of first-degree murder and sentenced to imprisonment (Imperial College Reporter, 1996; Nature, 1997).

Runcorn and mantle convection. Not long after his move to Cambridge in 1956, Keith Runcorn became interested in mantle convection as a driving mechanism for continental drift. In two papers in Nature in 1962 (Runcorn, 1962a,b), he established the theoretical arguments for subsolidus creep in the mantle under buoyancy stresses and made a quantitative prediction of convective velocities similar (although perhaps unbeknownst to him) to those of Hales and Pekeris. He also believed that a growing core influenced the onset of convective-driven drift, an idea that never found much traction.

Runcorn also proposed the then radical idea that the long-wavelength geoid was not frozen into the Earth but due to mantle convection (Runcorn, 1963). While Runcorn's analysis of the geoid was not as sophisticated as modern analysis (see **Chapters 7.02**, **7.04**, and **7.08**), it did lead him to infer that spreading centers were not sites of deep active upwellings, as had been inferred by Holmes and Hess, but were better explained by broad upwellings, which is a more modern and plausible view of mantle flow (see **Chapter 7.08**). This observation also prompted Runcorn to have little faith in the significance of the ridge-push force (Girdler, 1998; Runcorn, 1974). In many ways, Keith Runcorn's views of mantle convection were well ahead of their time and have largely been borne out by the last several decades of analysis.

Additional reading and historical sources for this brief history are Collinson (1998) and Girdler (1998).

7.01.2.8 Mantle Convection Theory in the Last 40 Years

The origin of modern mantle convection theory was contemporaneous not only with the birth (or rebirth) of the theory of plate tectonics but also with the advent of the study of nonlinear convection. Up till the early 1960s, the study of linear convection, that is, the onset of convective instability from infinitesimal perturbations (see **Chapters 7.02** and **7.04**), was well studied across several fields, as summarized in the classic treatise by Subrahmanyan Chandrasekhar (1910–95; Nobel Prize 1983) (Chandrasekhar, 1961). However, the mid-1960s witnessed increased activity in experimental studies of convection and in theories of finite-amplitude or nonlinear convection (see Manneville, 2006). Well-controlled experiments that documented transitions in convective state (e.g., patterns) (see **Chapter 7.03**) provided inspiration and testing for various studies of nonlinear convection, leading to several seminal theoretical approaches such as nonlinear perturbation theory and matched-asymptotic analysis (see **Chapter 7.04**).

Along with the plate tectonics revolution and the renewed study of the physics of convection, mantle convection theory made rapid progress and growth and achieved a rather mature level in the late 1960s and early 1970s (e.g., Oxburgh and Turcotte, 1978; Turcotte and Oxburgh, 1972) through the work of various investigators who had migrated in from physics and engineering, most notably Dan McKenzie, Jason Morgan, Donald Turcotte, and Gerald Schubert, the former two also having made major contributions to formulating the modern working theory of plate tectonics (see **Chapter 6.02**). Throughout the late 1960s and 1970s, many of the fundamental problems of convection in a solid-state mantle had been identified, although not necessarily solved. This included the fact that mantle convection is occurring in a fluid with a complex variable rheology, as well as through multiple solid–solid phase transformations, most notably at the top and bottom of the Earth's transition zone at depths of 410 and 660 km, respectively (see **Chapters 1.21**, **2.08**, **7.02** and **7.08**). This work set the stage for the explosion in the field of mantle dynamics that occurred during rapid improvements in computing power and numerical methods (see **Chapter 7.05**). In particular, many seminal contributions in the numerical analysis of mantle convection were made by Ulrich Christensen and colleagues during the 1980s and 1990s. With the combination of laboratory, theoretical, and numerical analyses, along with improved observations coming from, in particular, seismology, isotope geochemistry, and mineral physics, our picture of mantle convection has progressed in the last 20–30 years to highly sophisticated levels of complexity and realism, although much still remains to be understood.

7.01.3 Observations and Evidence for Mantle Convection

Many of the observations that are relevant to mantle dynamics are covered in other volumes of this treatise, as well as within this volume. As discussed in **Section 7.01.2**, the modern theory of mantle convection was motivated as the driving mechanism for continental drift and plate tectonics. The observations that in themselves inspired the resurgence of the mobile-surface

theory of plate tectonics were largely from paleomagnetic studies showing relative continental motion and seafloor spreading (see Sections 7.01.2.6 and 125.2.7; Chapter 6.02). Global seismicity also gave clear delineation of the structure of plates and locations of plate boundaries (Chapter 4.18) and outlines of subducting slabs along zones of deep earthquakes or Wadati–Benioff zones (Chapter 4.13). With the advent of the space program and satellite measurements came accurate geodetic measurements of sea-surface height and hence global models of the Earth's geoid and gravity field (Chapter 3.02), which provide important constraints about the density structure of the mantle associated with convection (Chapter 1.27).

Mantle geochemistry and petrology also provided important constraints on mantle dynamics through the analysis of magma reaching the surface from the mantle at mid-ocean ridges, ocean islands, and large igneous provinces and at subduction-related arcs. Isotopic and petrologic analyses of melting and melt fractionation of trace elements gave important information regarding the depth of melting beneath ridges and hot spots. However, the disparity between the concentration of incompatible elements in mid-ocean ridge basalts (MORBs) and ocean island basalts (OIBs), in addition to a host of other geochemical arguments involving, for example, noble gas isotopes, has been one of the driving motivations for inferences about the preservation of isolated reservoirs (e.g., layering) in the mantle. These geochemical observations, however, seem to conflict with geophysical evidence for whole-mantle stirring by sinking slabs (Grand et al., 1997; van der Hilst et al., 1997), and this has engendered a long-standing debate about the structure and nature of mantle convection (see Section 7.01.6.2, Chapter 2.04).

Our understanding of mantle dynamics also draws from observations of other planets. The lack of cratering record on Venus argues for massive resurfacing events, and giant volcanoes on Mars are evidence of extensive and deep magmatism (see Chapters 10.08 and 10.10). However, that neither of our neighboring terrestrial planets appears to have at least present-day plate tectonics continues to be one of the cornerstones in the argument that plate tectonics requires liquid water (Tozer, 1985; see Chapter 7.07).

7.01.4 Mantle Properties

The study of mantle convection has a boundless appetite for information on the properties of the convecting medium. Indeed, what caused the theory of continental drift to be marginalized for decades was a material property argument by Sir Harold Jeffreys that the Earth was too strong to permit movement of continents through ocean crust. Later measurements of mantle viscosity by postglacial rebound were an important key element in recognizing that the mantle is fluid on long time scales (see Chapter 7.02).

Fluid dynamics is rife with dimensionless numbers, and one of the most important such numbers in the study of convection is the Rayleigh number (see Chapter 7.02). The Rayleigh number defines the vigor of convection in terms of the competition between gravitationally induced thermal buoyancy, which acts to drive convective flow, and the dissipative or resistive effects of both fluid viscosity, which retards convective motion, and thermal diffusion, which acts to diminish thermal anomalies. The Rayleigh number is written generically as

$$Ra = \frac{\rho g \alpha \Delta T d^3}{\kappa \mu} \quad [1]$$

where ρ is the density, g is the gravitational acceleration, α is the volumetric thermal expansivity, ΔT is the typical temperature contrast from the hottest to coldest parts of the fluid layer, d is the dimension of the layer such as the layer thickness, κ is the thermal diffusivity, and μ is the dynamic viscosity (again, see Chapter 7.02).

Estimate of the mantle Rayleigh number by itself requires knowledge of the material responses to various inputs such as heat and stress. The response to heat input involves heat capacity, thermal expansivity α, and heat conduction or thermal diffusion (κ) (see Chapter 7.02, 2.08, and 2.23). The density structure inferred from high-pressure and high-temperature experiments (Chapter 2.08) and seismology (Chapters 1.01 and 1.21) constrain how mantle density ρ responds to pressure changes as upwellings and downwellings traverse the mantle, undergoing simple compression or decompression and solid–solid phase transitions; both effects can have either stabilizing or destabilizing effects on mantle currents (Chapters 7.02 and 7.09).

One of most important factors within the Rayleigh number and in the overall study of mantle convection is viscosity μ. That the mantle is viscous at all was one of the key elements in determining the viability of the mantle convection hypothesis. That continents were inferred to be in isostatic balance implied that they are floating in a fluid mantle; but isostatic equilibrium does not indicate how fluid the mantle is since it gives no information about how long it takes for the floating continents to reach an isostatic state. However, measurements of this approach to isostasy could be taken by examining postglacial rebound, that is, the uplift of high-latitude continental masses such as Scandinavia and Canada, following the melting of the glacial ice caps after the end of the last ice age (Haskell, 1937; see Chapter 3.07). From these analyses came one of the most crucial and well-known material properties: the average viscosity of the mantle of $\mu = 10^{21}$ Pa s. Today, the analysis of the mantle's response to changing loads (e.g., melting ice caps or a decrease in the Earth's rotation rate) is done through an increasingly sophisticated combination of geodetic satellite and field analyses, which further refine the viscosity structure of the mantle (Chapters 1.27, 3.07, 7.02, and 7.04).

The viscosity of the mantle is so large that it is called a slowly moving or creeping fluid and thus does not suffer the complexities of classical turbulence (Chapter 7.02). However, one of the greatest of all complexities in mantle dynamics is associated with the various exotic rheological behaviors of mantle rocks, which are almost exclusively inferred from laboratory experiments (Chapter 2.18). Mantle viscosity is well known to be a strong function of temperature, and the dramatic increase in viscosity toward the surface leads to various conundrums about how plate tectonics forms or functions at all and/or how subduction zones can ever initiate from such a cold strong lithosphere (Chapters 2.18, 7.07, and 7.09). The mantle's rheology is also complicated by the various deformation mechanisms it can assume, prevalently diffusion

creep at 'low' stress and dislocation creep at higher stress, although other creep and slip mechanisms are also possible (Chapters 2.18, 7.02, and 7.10). In diffusion creep, viscosity is a function of mineral grain size, and this effect can induce dramatic changes if grain growth or grain reduction mechanisms exist. In dislocation creep, viscosity is non-Newtonian and is a function of stress itself, thereby undergoing pseudoplastic behavior in which the material softens the faster it is deformed. These and many more complexities (see Chapters 7.02 and 7.07) continue to keep mantle convection a rich field.

7.01.5 Questions About Mantle Convection We Have Probably Answered

7.01.5.1 Does the Mantle Convect?

The existence of this entire volume as well as an enormous body of literature on mantle convection seems to obviate this question. However, we should understand that the notion that the mantle convects was not entirely accepted less than 50 years ago. The physical requirements or conditions for a fluid mantle to convect (i.e., a sufficiently high Rayleigh number) could be inferred from material property measurements; these indeed imply that the Rayleigh number is perhaps a million times what is needed to just barely convect at all and thus should be convecting vigorously (Chapter 7.02).

But the direct observation of a convecting mantle is most closely linked to those that verify plate tectonics and to seismic imaging of the Earth's interior. Plate spreading and subduction, and the 'creation' and 'destruction' cycle of lithosphere that they represent, demand vertical transfer from the surface into the mantle and vice versa. That heat flow and bathymetry measurements show that the lithosphere going from a hot ridge to a cold trench is evidence not only that ridges and trenches are the expression of upwellings and downwellings, respectively, but also that the cooling lithosphere is nothing more than a convective thermal boundary layer (Chapters 7.02, 7.04, and 7.07). Lastly, seismology has given us not only the deep-earthquake trace of a the cold sinking slab along the Wadati–Benioff zone (Chapter 4.13) but also tomographic images of these same slabs sinking deep into the mantle (Grand et al., 1997; van der Hilst et al., 1997). There remains little if any doubt that the mantle convects, but it is important to remember what the first-order evidence is for this conclusion, especially considering that the mantle is more inaccessible to direct observation than distant galaxies.

7.01.5.2 Is the Mantle Layered at 660 km?

The average structure of the mantle in terms of density and elastic properties was shown by seismological studies (augmented by mineral physics) to contain discontinuities, most notably at about 410 and 660 km depths (Chapters 1.01, 1.21, and 2.08), the latter one being considerably more distinct. The presence of a strong discontinuity in density at 660 km depth suggested that the lower mantle was a denser and perhaps isolated and sluggish layer convecting separately from the upper mantle. This layered convection argument was reinforced by observations that deep earthquakes along subducting slabs cease at around 700 km depth (Chapter 4.13) and that some tomographic images show slabs stalling at this depth. This view also fit well with the geochemical inferences that MORBs and ocean island basalts must be coming from different layers (the former from the upper mantle above 660 km and the latter from the lower mantle below 660 km; see also Section 7.01.6.2 and Chapter 7.10). The convergence of seismology and geochemistry toward a single model of a mantle layered into compositionally distinct regions was indeed a compelling argument. The prospect of mantle convection existing in two layers separated at 660 km depth was a prevalent theme in the study of mantle convection for several decades starting in the late 1960s.

However, high-pressure mineral physics experiments indicated that mantle discontinuities are most likely associated with solid–solid phase transitions, not compositional changes. The major upper mantle component olivine was shown to undergo a change to a spinel structure called wadsleyite at 410 km depth; wadsleyite itself undergoes a less dramatic transition to a ringwoodite at around 510 km and, then, at 660 km depth, ringwoodite changes to a combination of perovskite and magnesiowüstite (Chapters 2.08 and 7.02). Studies of convection in the presence of such phase changes indicated that they might impede convection temporarily but not indefinitely (Chapters 7.02, 7.08, and 7.09); and indeed, seismic tomographic studies using body waves showed that many slabs do indeed penetrate this boundary and sink well into the lower mantle (Grand et al., 1997; van der Hilst et al., 1997).

In the end, the predominant evidence points to the mantle not being layered with an impermeable boundary at 660 km depth, although this boundary does appear to inhibit or temporarily stall vertical transport of slabs. This conclusion, however, leads to other unsolved conundrums, especially with regard to explaining geochemical observations, which has thus inspired several variants of deep mantle layering and ways of isolating reservoirs or chemical components (Chapter 7.11).

7.01.5.3 What Are the Driving Forces of Tectonic Plates?

Upon the widespread acceptance of the plate tectonic model, considerable effort was put forward to make direct estimates of the plate-driving forces. This effort was perhaps best represented by the seminal work of Forsyth and Uyeda (1975), although considerable work has followed since then (e.g., Jurdy and Stefanick, 1991). Several plate forces have been coined, most notably the two driving forces of ridge push and slab pull. Which of these two forces is dominant was the subject of some debate (Hager and O'Connell, 1981; Jurdy and Stefanick, 1991). However, perhaps, the most compelling evidence for slab pull is the profound correlation, shown in the original paper of Forsyth and Uyeda (1975), between the fraction of convergent (trench) boundary and plate velocity (Figure 9); essentially, no other meaningful correlation was shown between any other type of boundary and plate velocity. This correlation showed that the fastest plates (of order 10 cm year^{-1}) have the most amount of slab connected to them, while the slowest plates (of order 1 cm year^{-1}) have little or no slab connected to them. This demonstrated rather conclusively that slab pull is the driving force of plate tectonics, because quite simply, for a plate to move, it needs a slab (see Chapter 7.09). What is also important about this inference from a mantle convection perspective is that it fits very well with the modern picture of cold downwelling or slab-dominated mantle circulation. That is, the Earth's mantle is primarily

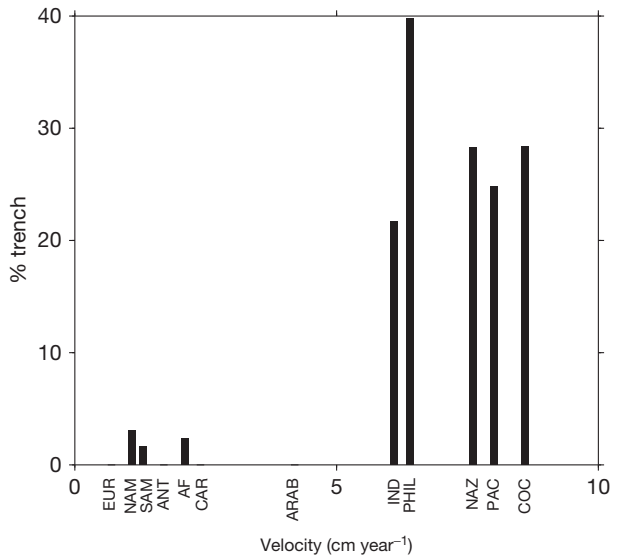


Figure 9 A histogram of percent trench length (percent that a plate boundary is composed of subduction zone) versus plate velocity for the major plates (initials indicated). (Initials on the abscissa are for European (EUR), N. American (NAM), S. American (SAM), Antarctic (ANT), African (AF), Caribbean (CAR), Arabian (ARAB), Indo-Australian (IND), Philippine (PHI), Nazca (NAZ), Pacific (PAC), and Cocos (COC) plates.) Adapted from Forsyth D and Uyeda S (1975) On the relative importance of the driving forces of plate motion. *Geophysical Journal of the Royal Astronomical Society* 43: 163–200; Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121, with permission from Elsevier.

driven by the surface cooling of a mantle that is more or less uniformly heated by radioactive decay of uranium, thorium, and (at one time) potassium and is also losing primordial or fossil heat (Chapter 7.06); such a configuration of distributed heat production and surface heat loss typically leads to convection dominated by cold downwelling currents, which are synonymous with subducting slabs (see Chapters 7.02, 7.07, and 7.09). Most importantly of all, the estimate of plate-driving forces is thus completely reconcilable with and even subsumed by the theory of mantle convection. Moreover, this inference leads to the important conclusion that the tectonic plates are not so much driven by convection; they are convection. The plates are cooling thermal boundary layers that are both driven by and become slabs, which are in themselves convective downwellings; the plates are thus convection (see further discussion in Chapter 7.07).

7.01.6 Major Unsolved Issues in Mantle Dynamics

7.01.6.1 Energy Sources for Mantle Convection

The discovery of radioactive elements at the turn of the nineteenth century was a key discovery in many regards, including providing evidence that the Earth has internal sources of heat. This was also the key argument to refute Lord Kelvin's estimate for the age of the Earth since he assumed that it was cooling freely (without heat sources) from an initially molten state (see Section 7.01.2.4). Early estimates of the concentration of radioactive elements inside the Earth are based on heat flow and geochemical measurements of crustal rocks, in conjunction with cooling models (Chapter 7.05); these arguments tended to point toward a high enough concentration of radioactive elements in the mantle to account for as much as 70–80% of the heat flow out of the Earth to be due to radiogenic heating (Schubert et al., 2001). This satisfied the condition that the Earth has been cooling relatively slowly over its 4.5 Gy lifetime. However, recent estimates of radioactive element abundances from the study of chondrites (thought to be representative of planetary building blocks) possibly suggest somewhat less radiogenic heating (see Chapter 7.06). If radiogenic heating is small, then more of the Earth's heat flow is from loss of fossil heat (as with Lord Kelvin's assumptions; see also England et al., 2007); this would demand rapid cooling from a recently excessively hot or even molten state, unless the physical process of convection was itself somehow very different in the past in order for the mantle to retain its heat (see Chapters 7.06 and 9.06). Alternatively, if chondritic estimates of these radiogenic sources are very wrong, then it implies significant problems with the chondritic model for planetary composition and thus for our understanding of the Earth's formation. Either possibility leads to many intriguing directions for future inquiry.

7.01.6.2 Is the Mantle Well Mixed, Layered, Piled, or Plum Pudding?

As mentioned in the preceding text in Sections 7.01.3 and 125.5.2, there are various lines of geochemical evidence – ranging from the disparity between trace element abundances in MORBs and those in ocean island basalts, sources and reservoirs of noble gases, source and origin of continental crust, etc. – that suggest that the mantle has isolated reservoirs, such as distinct layers. Although layering of the mantle at 660 km depth has probably been eliminated, the motivation still persists to reconcile the geochemical inference of an unmixed mantle with the geophysical evidence of a well-stirred mantle. The problem remains unsolved, although various models and solutions have been proposed, in particular deep layering, or mechanisms for keeping the mantle poorly mixed (the 'plum-pudding' model). In recent years, deep-layering models have evolved into 'pile' models wherein the dense and long-lived abyssal layer is noncontiguous and gathered into at least two stable mounds or piles at the core–mantle boundary; seismic evidence for these piles has grown and the mantle-pile model has become an active area of research and debate (see Chapter 7.11). This issue comprises the major thrust of Chapter 7.11 and is touched upon in Chapter 7.10, as well as in other volumes, notably Chapter 2.04.

7.01.6.3 Are There Mantle Plumes?

Convective plumes are relatively narrow cylindrical upwellings typical of convection and especially in fluids with strongly temperature-dependent viscosity, such as mantle rocks. The existence of plumes in the mantle was proposed by Jason Morgan (see Chapter 7.10) to explain anomalous intraplate volcanism such as at Hawaii. That hot spots appeared to be more or less immobile relative to plates suggested a deep origin, and it has often been supposed that plumes emanate from the most obvious heated boundary in the mantle, the core–mantle boundary. Direct observation of plumes by seismic methods has achieved some success in recent years in identifying hot mantle anomalies below, for example, Hawaii (Wolfe et al.,

2009), as well as by newer techniques that still remain controversial (Montelli et al., 2004). Although evidence for the existence of plumes has grown in the last decade, they remain the subject of ongoing debate. The subject of plumes and melting anomalies is discussed in **Chapter 7.10**.

7.01.6.4 Origin and Cause of Plate Tectonics

As noted already, the modern theory of mantle convection was largely motivated to provide a driving mechanism for plate tectonics. We also argued in the preceding text (**Section 7.01.5.3**) that plate tectonics is indeed convection. However, there still remains no unified theory of mantle dynamics and plate tectonics, wherein plate tectonics arises naturally and self-consistently from mantle convection. There are some aspects of plate tectonics that are reasonably well explained by convective theory, in particular the existence of cold planar downwellings akin to subducting slabs (**Chapter 7.09**). But still, many first-order questions remain. Just with regard to subduction itself, there is still no widely accepted theory of how a subduction zone and sinking slab initiate from a thick cold and, by all appearances, immobile lithosphere and sink asymmetrically; that is, only one side of a convergent zone descends (see **Chapters 7.02**, **7.07**, and **7.09**).

But many other issues remain in the problem of 'plate generation.' Although mid-ocean ridges are associated with upwelling from the mantle, all evidence points to the upwelling being shallow and the spreading being passive; that is, ridges are pulled apart by slabs at a distance, rather than pried apart by a deep upwelling (**Chapter 7.08**). How such passive upwelling occurs in convection is still not universally understood. Another long-standing enigma of the platelike form of mantle convection is the existence of toroidal motion, which involves strike-slip shear and spin of plates (Dumoulin et al., 1998; Hager and O'Connell, 1979; O'Connell et al., 1991). Toroidal motion is enigmatic because it is not typical of convective flow; that is, it is not directly driven by thermal buoyancy forces and likewise does not transport heat and thereby serves no apparent purpose (see **Chapters 7.02**, **7.04**, and **7.07**).

A generalizable theory of how plate tectonics is generated by mantle convection is crucial to understanding not only how Earth works but also how planets in other solar systems work, that is, whether they are more likely to look like Earth, Venus, Mars, or something unexpected. In particular, since plate tectonics is believed to stabilize the Earth's clement climate for hundreds of millions of years, it is considered a necessary condition for habitability and thus an active topic of debate in the study of extrasolar terrestrial planets. Much still remains to be learned in the fundamental problem of plate generation and is the main topic of **Chapter 7.07**, as well as being discussed throughout this treatise (see **Chapter 9.06**).

7.01.7 Burgeoning and Future Problems in Mantle Dynamics

7.01.7.1 Volatile Circulation

The interaction of the ocean and atmosphere with the mantle has in fact been an important and fertile field of study for the last few decades, although it has largely been the province of mantle petrology and geochemistry. However, questions of how much the mantle entrains, returns, and stores various important volatiles such as water and carbon dioxide remain important and largely unanswered (Dasgupta and Hirschmann, 2010; Hirschmann, 2006; Karato, 2003; Williams and Hemley, 2001). The relevance of the problem is manifold but can be summarized perhaps in two themes: First, the ingestion of volatiles by the mantle affects its convective circulation and thus both the thermal and chemical evolution of the mantle, mainly because of the rheological effects of volatiles (which tend to weaken rocks) and their tendency to facilitate melting and hence chemical and isotopic fractionation. Second, how the mantle ingests, stores, and releases various volatiles controls the evolution of the oceans and atmosphere, in both their size (or mass) and composition (since different volatiles are likely to be entrained and stored differently).

The flux of volatiles into the mantle is primarily through subduction zones (Hirschmann, 2006; Peacock, 1990). Crustal rocks entering a subduction zone are necessarily hydrated by virtue of being submarine, leading to, for example, hydrated sediments and serpentinized olivine. Whether the lithosphere as a whole is hydrated is questionable since it tends to be dried during the process of mantle melting and formation of crust at ridges. Absorption of carbon dioxide is through weathering and carbonation reactions with silicates to produce carbonates (Falkowski et al., 2000), which are then entrained by subduction. However, the quantification of volatile entrainment through subduction zones is problematic for many reasons. In the case of water, it is difficult to make an accurate estimate of how much gets carried by the slab, how much continues to be carried down after devolatilization through arc magmatism, and which hydrous silicate phases in the slab are capable of carrying water to significant mantle depths (Hirschmann, 2006; Williams and Hemley, 2001; see also **Chapter 2.04**). Carbonates are likely more stable than water at depth, thus leading to greater mantle ingestion of carbon, depending on mantle temperatures; indeed, carbon entrainment and degassing were likely very different in the past than currently (Dasgupta and Hirschmann, 2010).

Storage of volatiles in the mantle is also the subject of much debate. Although solid rocks cannot absorb volatiles in great concentrations, the mass of the mantle is so large that it could conceivably hold several to tens of world ocean masses. Moreover, the region of the mantle between the phase transitions at 410 and 660 km depths – called the transition zone – is known for having anomalous solubility of at least water, and while it is only a tenth of the thickness of the mantle, it could hold much more water than either the upper mantle (above 410 km) or lower mantle (below 660 km) combined. The transfer of anomalous water from the transition zone through convection could possibly cause deep melting, as has been proposed since the early 1990s (e.g., Inoue, 1994; Inoue and Sawamoto, 1992); this melting might have an effect on trace element circulation and lead to the appearance of layering in whole-mantle convection (Bercovici and Karato, 2003; Karato et al., 2006). How much volatile mass is eventually returned to the oceans and atmospheres through volcanism is perhaps better constrained since most nonarc volcanic output occurs at

mid-ocean ridges whose basalts are relatively dry. The ingestion of water by slabs and meager water return at ridges suggests that the mantle is on the whole absorbing oceans, although it is equally possible that little water is taken into the mantle beyond arc volcanism (Dixon et al., 2002; see Chapter 1.10; and Hirschmann, 2006). The balance of volatiles between the oceans and atmosphere is an ongoing debate since it relies on various feedback mechanisms, some of which are still not well articulated. Volatiles will tend to reduce mantle viscosity and enhance convective vigor, but whether this enhances ingestion or output of water is not entirely known and similar models can lead to very different conclusions, such as complete mantle degassing in a brief time (McGovern and Schubert, 1989), or drainage of the oceans into the mantle over billions of years (Bounama et al., 2001). However, evidence that ocean masses have remained more or less constant over billions of years (see Hirschmann, 2006) implies a steady-state exchange of the mantle with the oceans and atmosphere that either has a trivial explanation (i.e., there is no exchange) or requires a self-regulating feedback mechanism (see Korenaga, 2008). Such mechanisms might involve the ocean–mantle contact area that governs the size (depth and breadth) of the world's oceans and thus the size of ocean basins, plate boundary lengths, and plate sizes. To simply state that the subduction and ridges must ingest and eject an equal amount of water is not an explanation since the mechanisms of absorption at subduction zones and release at ridges are vastly different, and it would be fortuitous if they could balance each other for any given plate or ocean size configuration.

7.01.7.2 Mantle Convection: Water and Life

The presence of liquid water on Earth is likely a necessary condition for plate tectonics and an obvious necessary condition for the existence of life as we know it (see Chapter 10.14). A major question still remains as to whether all three are linked, that is, that plate tectonics has allowed liquid water and life or possibly even whether life has influenced plate tectonics. The volcanic return of subducted carbon dioxide, stored as carbonates in oceans sediments, likely sustains a greenhouse state that keeps the surface temperature sufficiently high for water to remain liquid and hence permit plate tectonics. Moreover, the absorption of carbon dioxide by weathering and carbonation of minerals that are continuously extruded or exhumed by plate tectonics and convection prohibits the buildup of CO_2 and a possible runaway greenhouse; indeed, the negative feedback associated with weathering (e.g., where higher atmospheric CO_2 leads to warmer temperatures, hence more precipitation and weathering, and thus faster drawdown of CO_2) is thought to stabilize climate over geologic time scales (Berner, 2004; Walker et al., 1981) and thus allow the evolution of complex life (Ward and Brownlee, 2000). However, the rise of life itself is also associated with major climatic effects, including carbon sinks in the form of biomass and kerogens (Falkowski et al., 2000) as well as oxidation of the atmosphere that could have removed potent greenhouse gases like methane (Kasting and Howard, 2006). Thus, life's role in stabilizing climate and sustaining liquid water may have influenced the evolution of plate tectonics also.

7.01.8 Summary and Context of the Rest of This Volume

This volume on mantle dynamics is designed to follow two themes: (1) how is mantle convection studied and (2) what do we understand about mantle dynamics to date. The first four chapters following this overview are thus concerned with pedagogical reviews of the physics of mantle convection (Chapter 7.02), laboratory studies of the fluid dynamics of convection relevant to the mantle (Chapter 7.03), theoretical analysis of mantle dynamics (Chapter 7.04), and numerical analysis of mantle convection (Chapter 7.05). The subsequent chapters concentrate on leading issues of mantle convection itself, which include the energy budget of the mantle (Chapter 7.06); the generation of plate tectonics from mantle convection (Chapter 7.07); the upper mantle and lithosphere in and near the spreading center (mid-ocean ridge) environment (Chapter 7.08); the dynamics of subducting slabs (Chapter 7.09); hot spots, melting anomalies, and mantle plumes (Chapter 7.10); the dynamics of the deep lower mantle and the core–mantle boundary (Chapter 7.11); and lastly, geochemical mantle dynamics and mixing (Chapter 7.12).

The physics of mantle convection is extensively multidisciplinary since it involves not only fluid mechanics but also gravitational potential theory (to understand not only self-gravitation of our massive convecting medium but also how the shape of the geoid or sea surface is affected by and informs us about convection), seismology and mineral physics (to understand the thermodynamic state and properties of the mantle such as density, thermal expansivity, and solid–solid phase transitions), material science (to understand rheology and deformation mechanisms and transport phenomena), and geochemistry, petrology, and complex multiphase, multicomponent flows (to understand chemical mixing and mantle melting). Chapter 7.02 begins by elucidating the basic physics of convection universal to any fluid system and then proceeds to discuss the complexities associated with the mantle, including, compressibility, phase changes, viscoelasticity, and silicate rheology. Many of the future research topics involving mantle complexities cannot be described with single-component and single-phase fluid mechanics; thus, some care is given to develop an introduction to complex fluids entailing multicomponent systems undergoing mixing and chemical transport, as well as deformable multiphase media.

As discussed in Section 7.01.2.3, the initiation of the modern study of convection is attributed to the first systematic experimental studies of Henri Bénard. Thus, Chapter 7.03 involves a survey of laboratory methods for studying convection and for approaching the particular complexities of mantle convection itself. Convection in the mantle is, of all geophysical flows, perhaps most easily scaled to laboratory conditions since the ratio of viscosity to the cube of layer thickness is easily preserved between the mantle and the laboratory model. Moreover, laboratory models do not suffer the assumptions made in physical theory to obtain a closed system, the simplifying approximations of analytic theory to obtain a mathematical solution or the limitations of numerical resolution in computer models; thus, the need for studying convection in real materials is paramount. However, the mantle is an exotic fluid in that it is undoubtedly chemically

inhomogeneous, has phase changes, is heated and cooled internally, and has rheology sensitive to various state variables such as temperature, pressure, and stress or grain size, and these effects are all difficult to reproduce in the laboratory and with available laboratory fluids. Thus, progress in laboratory models is vital for both exploring new physics and testing theories but remains an extremely challenging field. Chapter 7.03 thus not only discusses the methods for creating and observing convection experiments but also reviews the considerable progress in incorporating the complexities of the mantle itself into these experiments.

The birth of convection as a field of physics is also associated with Lord Rayleigh's seminal theoretical work on convective instability (Section 7.01.2.2). In the last 40–50 years, the theoretical analysis of convection in both the linear and especially nonlinear regimes has burgeoned, not only because of its relevance to planetary and stellar atmospheres and interiors but also because convection is the classic paradigm of a nonlinear dynamic system undergoing chaotic behavior and self-organization (e.g., Nicolis, 1995). Thus, Chapter 7.04 surveys the wealth of theoretical analyses on convection itself and on individual theories relating to features of convective circulation. Thus, the initiation of mantle plumes is examined through stability (Rayleigh–Taylor) theory and fully developed plumes and mantle diapirs through simpler fluid dynamic models. Likewise, subducting slabs are examined through, for example, the theory of bending viscous sheets. Finally, convection is studied holistically through a review of weakly nonlinear perturbation theories, which predict three-dimensional patterns at convective onset, to matched-asymptotic boundary layer theories for strongly supercritical convection.

Progress in the study of mantle dynamics in the last 20 years has been driven by the rapid increase in computational power as well as ever-growing sophistication in numerical models. Numerical methods are invariably the most versatile of all methods of study since they can examine strongly nonlinear convection without excessive simplifying assumptions and can incorporate any relevant physics that can at least be articulated mathematically. Thus, Chapter 7.04 reviews the numerical analysis of mantle dynamics by surveying the leading methods that are employed today. Classic methods of numerical modeling include finite-difference, finite-volume, and spectral methods, all of which are outlined in Chapter 7.04; but perhaps, the most powerful and versatile method is finite elements and this is given special attention. Some classic examples of how numerical analysis is used to attack key problems are also discussed; these include the omnipresent mantle complexities of thermochemical convection, phase changes, and non-Newtonian rheology.

With the major tools of studying mantle convection surveyed in detail, the volume progresses to issues specific to the mantle itself. An obvious and key issue of mantle dynamics is the energy source for convection, which remains an active and at times controversial issue, and this is reviewed in Chapter 7.06. The energy budget of the Earth relies on various important quantities that are unfortunately not easily constrained. The loss of heat through the Earth's surface is a first-order observation but is problematic in that the measurements are difficult to make (e.g., measurements of conduction through the lithosphere are easily contaminated by the effects of hydrothermal circulation), and global coverage of heat flow is difficult given the large variation in both crustal thickness and properties between and within oceans and continents. Secondly, the estimate of the heat sources inside the Earth is also difficult since it involves understanding the composition of the bulk Earth, which involves measurements of radioactive element concentrations from various sources such as continental crust, oceanic basalts, and chondritic meteorites, but of course never directly from the mantle itself. Finally, basic understanding of convection from theoretical, numerical, and analytic models (Chapters 7.03–7.05) allows construction of thermal histories of how the Earth cools and has evolved under the action of convection.

Chapters 7.07–7.12 concentrate on some of the key individual features of convective circulation itself. Chapter 7.07 explores one of the grand challenge problems of geodynamics as discussed already in Section 125.6.4, that is, the relation between plate tectonics and mantle dynamics and in particular how and why plates are generated from a cooling convecting mantle on Earth (but not apparently other terrestrial planets). It is widely recognized that the plates and lithosphere are the dominant convective thermal boundary layer of the mantle as it is cooled from above. Indeed, some aspects of simple convection theory explain tectonic features reasonably well (e.g., the rough shape of cold downwellings and slabs). However, a host of other features of plate boundaries, such as narrow passive ridges, strike-slip (i.e., toroidal) motion, and asymmetric subduction, are not well explained and require extra complexity. Such complexity usually entails additional physics of rock deformation that is not typical of viscous fluids, such as highly non-Newtonian rheology and/or nonlinear weakening or 'damage' mechanisms that provide a feedback between flow and strength. That plate boundaries remain weak when dormant and can be reinherited to form new active boundaries is another important feature requiring weakening to have memory or accumulated history. Finally, a general theory of how plates are generated is important for gleaning the origin of plate tectonics in the Archean and the conditions for plate tectonics on other planets.

Chapter 7.08 treats the problem of convection in the upper mantle and its interaction with mid-ocean ridges and the oceanic lithosphere. How the oceanic lithosphere forms following complex magmatic processes at ridges can play a key role in the entire convective cycle, since, for example, melting at ridges can cause dehydration and strengthening of the lithosphere. How this same lithosphere transmits heat and thickens as it moves away from mid-ocean ridges is an important feature of mantle convection. However, small-scale convection (as well as hot spot activity) in the sublithospheric asthenosphere can have a profound effect on how heat is transferred into the lithosphere. Moreover, plate motions can affect the structure of this small-scale convection, typically by aligning convective rolls with plate motion. Similar to small-scale convection is the convective instability of the lithosphere itself, which can lead to delamination and further change of lithospheric structure and cooling.

Chapter 7.09 continues with the plate and lithosphere as it enters subduction zones and becomes what is clearly the driving force for mantle convection, that is, subducting slabs. The chapter first reviews the basic mechanism and energy release of slab descent. It then proceeds to examine evidence for slab structure from seismological observations, which can then be

used to constrain dynamic models of sinking slabs as they descend through the subduction zone wedge environment and through mantle phase transformations. Phase changes have a particularly important effect on slabs because not only they can impede slab transfer into the lower mantle but also the slow kinetics inside cold slabs might cause metastable phases to exist, which potentially cause fine-scale buoyant stresses that could influence slab deformation and seismicity. One of the primary constraints on slab-related convection and mantle structure involves the geoid and topographic signature of subduction zones; in particular, the geoid over subduction zones is generally positive, which suggests that the cold positive mass anomaly of the slab is supported by a viscosity increase with depth rather than by a downward deflection of the surface. Lastly, as discussed in the preceding text, the volatile flux into the Earth is entirely coupled to subduction and slab dynamics, and this issue is covered throughout Chapter 7.09.

To contrast the discussion of cold slabs, Chapter 7.10 concerns anomalously hot mantle, melting anomalies, and the mantle's other primary convecting current, the elusive mantle plume. The chapter reviews first the observations relating to hot spots and melting, including discussion of age-progression volcanism (which has been the observational foundation of the stationary hot spot model), the topographic structure of hot spot swells (which suggest a larger mantle structure than just the volcanic edifice itself), the signature of large igneous provinces (thought to be associated with large starting plumes and/or the initiating of continental rifting), and the geochemical, petrologic, and seismological evidence for the depth of origin of hot spots. Next, the process of hot spot melting and volcanism, the formation of plume swells, and the fluid dynamics of deep mantle plumes are examined and evaluated.

Chapter 7.11 deals with the burgeoning field of mantle dynamics in the deep lower mantle and the core–mantle boundary. In recent years, mineral physics and seismology have suggested three major findings concerning the lowermost mantle, namely, the likely existence of regions of melt, called ultra-low velocity zones, the presence of two large chemical dense mounds or piles resting on the core–mantle boundary (perhaps analogous to large anti-continents), and the existence of a very high-pressure mineralogical postperovskite phase near the boundary. The melt zones are possibly coincident with the edges of the piles and may represent the remnants of the Earth's magma ocean (hence called the basal magma ocean). The deep mantle piles are possibly also the source for enriched mantle (i.e., they are the primitive abyssal layer) and are associated with two large regions of hot spot activity, that is, the South Pacific Superswell and the East African Superswell. Finally, the thickness and structure of the postperovskite layer help constrain mantle temperatures above the core–mantle boundary and hence the heat flow emanating out of the core, which might be larger than traditionally assumed. The influence of these new important features on mantle flow and evolution have been subjects of considerable research and debate in the last decade, especially as more information comes out about this most distant region of the mantle.

Involves the ongoing problem of interpreting observations of geochemical heterogeneity from the perspective of mantle convection theory. The chapter first reviews the observations and evidence for mantle heterogeneity, starting with its origin during early segregation of the mantle, crust, and core, to its present state. As discussed in the preceding text (Section 7.01.6.2), much of the evidence for current heterogeneity involves analysis of trace (incompatible) elements in mid-ocean ridge and OIBs, as well as budgets of noble gases, and these primary constraints are explained and reviewed as well. Following the survey of observations, the chapter examines models of mixing, trace element transport and layering in the mantle to address the fate, scale and isotopic signature of dispersed heterogeneities, and the stability of reservoirs of large-scale chemical heterogeneity. The problem of chemical evolution and heterogeneity remains one of the biggest unsolved problems in mantle dynamics, and thus, recent and future directions in this field are also discussed.

In the end, this volume of the *Treatise on Geophysics* is designed to give both a classical and state-of-the-art introduction to the methods and science of mantle dynamics and a survey of leading order problems (both solved and unsolved) and our present understanding of how the mantle works.

Acknowledgments

The author is indebted to Gerald Schubert for his comments, corrections, and, most of all, patience. Thanks are also due John Ferguson for the photograph of Anton Hales and Graham Pearson for the photograph of Arthur Holmes.

References

Ahrens R (1977) William Prout (1785–1850). A biographical sketch. *Journal of Nutrition* 107(1): 15–23.
Bénard H (1900) Les tourbillons cellulaires dans une nappe liquide. *Revue générale des Sciences pures et appliquées* 11: 1261–1271, 1309–1328.
Bénard H (1901) Les tourbillons cellulaires dans une nappe liquide transportant de la chaleur par convection en régime permanent. *Annales de chimie et de physique* 23: 62–144.
Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121.
Bercovici D and Karato S (2003) Whole mantle convection and the transition-zone water filter. *Nature* 425: 39–44.
Berg J, Acrivos A, and Boudart M (1966) Evaporative convection. *Advances in Chemical Engineering* 6: 61–123.
Berner R (2004) *The Phanerozoic Carbon Cycle*. Oxford: Oxford University Press.
Block M (1956) Surface tension as the cause for Bénard and surface deformation in a liquid film. *Nature* 178: 650–651.
Bounama C, Franck S, and von Bloh W (2001) The fate of earth's ocean. *Hydrology and Earth System Sciences* 5: 569–575.
Brown G (1999) *Scientist, Soldier, Statesman, Spy: Count Rumford – The Extraordinary Life of a Scientific Genius*. Thrupp, Stroud, Gloucestershire: Sutton Publishing Ltd.
Brown S (1967) *Benjamin Thompson – Count Rumford: Count Rumford on the Nature of Heat*. New York: Pergamon Press.
Buddington A (1973) Memorial to Harry Hammond Hess – 1906–1969. *Geological Society of America Memorials* 1: 18–26.
Bull A (1931) The convection current hypothesis of mountain building. *Geological Magazine* 58: 364–367.
Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. New York: Oxford University Press.
Collinson D (1998) The life and work of S. Keith Runcorn, F.R.S. *Physics and Chemistry of the Earth* 23: 697–702.
Dasgupta R and Hirschmann MM (2010) The deep carbon cycle and melting in earth's interior. *Earth and Planetary Science Letters* 298(1–2): 1–13.
Dixon JE, Leist L, Langmuir C, and Schilling J-G (2002) Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* 420: 385–389.

Dumoulin C, Bercovici D, and Wessel P (1998) A continuous plate-tectonic model using geophysical data to estimate plate margin widths, with a seismicity based example. *Geophysical Journal International* 133: 379–389.

Dunn D (1984) Harry Hammond Hess (1906–1969). *Essay for AGU Honors*. American Geophysical Union.

England P, Molnar P, and Richter F (2007) John Perry's neglected critique of Kelvin's age for the Earth: A missed opportunity in geodynamics. *GSA Today* 17(1): 4–9.

Falkowski P, Scholes RJ, Boyle E, et al. (2000) The global carbon cycle: A test of our knowledge of earth as a system. *Science* 290: 291–296.

Forsyth D and Uyeda S (1975) On the relative importance of the driving forces of plate motion. *Geophysical Journal of the Royal Astronomical Society* 43: 163–200.

Gilbert F (2004) Chaim Leib Pekeris, June 15, 1908–February 24, 1993. *Biographical Memoirs* 85: 217–231.

Gillis J (1995) Chaim L. Pekeris, 15 June 1908–25 February 1993. *Proceedings of the American Philosophical Society* 139: 179–181.

Girdler R (1998) From polar wander to dynamic planet: A tribute to Keith Runcorn. *Physics and Chemistry of the Earth* 23: 709–713.

Grand S, van der Hilst R, and Widiyantoro S (1997) Global seismic tomography: A snapshot of convection in the earth. *GSA Today* 7: 1–7.

Hager B and O'Connell R (1979) Kinematic models of large-scale flow in the Earth's mantle. *Journal of Geophysical Research* 84: 1031–1048.

Hager B and O'Connell R (1981) A simple global model of plate dynamics and mantle convection. *Journal of Geophysical Research* 86: 4843–4867.

Hales A (1936) Convection currents in the Earth. *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement* 3: 372–379.

Hallam A (1987) Alfred Wegener and the hypothesis of continental drift. In: *Scientific Genius and Creativity: Readings from 'Scientific American' (Reprinted from Scientific American, February, 1975)*, pp. 77–85. New York: W.H. Freeman & Co.

Haskell N (1937) The viscosity of the asthenosphere. *American Journal of Science* 33: 22–28.

Hess H (1962) History of ocean basins. In: Engeln A, James H, and Leonard B (eds.) *Petrologic Studies – A Volume in Honor of A.F. Buddington*, pp. 599–620. New York: Geological Society of America.

Hirschmann MM (2006) Water, melting, and the deep earth H_2O cycle. *Annual Review of Earth and Planetary Sciences* 34: 629–653.

Holmes A (1931) Radioactivity and earth movements. *Geological Society of Glasgow, Transactions* 18: 559–606.

Holmes A (1933) The thermal history of the Earth. *Journal of the Washington Academy of Sciences* 23: 169–195.

Imperial College Reporter (1996) Murder charge. *In Brief* (22).

Inoue T (1994) Effect of water on melting phase relations and melt composition in the system Mg_2SiO_4–$MgSiO_3$–H_2O up to 15 GPa. *Physics of the Earth and Planetary Interiors* 85: 237–263.

Inoue T and Sawamoto H (1992) High pressure melting of pyrolite under hydrous condition and its geophysical implications. In: Syono Y and Manghani M (eds.) *High-Pressure Research: Application to Earth and Planetary Sciences*, pp. 323–331. Washington, DC: American Geophysical Union.

Jeffreys H (1959) *The Earth, Its Origin, History and Physical Constitution*, 4th edn. Cambridge: Cambridge University Press.

Jurdy D and Stefanick M (1991) The forces driving the plates: Constraints from kinematics and stress observations. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 337: 127–138.

Karato S (2003) Mapping water content in the upper mantle. In: Eiler J (ed.) *Subduction Factory. AGU Monograph*, pp. 135–152. Washington, DC: American Geophysical Union.

Karato S, Bercovici D, Leahy G, Richard G, and Jing Z (2006) The transition zone water filter model for global material circulation: Where do we stand? In: Jacobsen S and van der Lee S (eds.) *Earth's Deep Water Cycle. AGU Monograph Series*, vol. 168, pp. 289–313. Washington, DC: American Geophysical Union.

Kasting J and Howard M (2006) Atmospheric composition and climate on the early Earth. *Philosophical Transactions of the Royal Society B* 361: 1733–1742.

Korenaga J (2008) Plate tectonics, flood basalts, and the evolution of Earth's oceans. *Terra Nova* 20: 419–439.

Lambeck K (2002) Professor Anton Hales – Terrestrial and planetary scientist. Interviews with Australian Scientists Program, Australian Academy of Science.

Leitch A (1978) *A Princeton Companion*. Princeton, NJ: Princeton University Press.

Lewis C (2000) *The Dating Game – One Man's Search for the Age of the Earth*. Cambridge: Cambridge University Press.

Lewis C (2002) Arthur Holmes: An ingenious geoscientist. *GSA Today* 12(3): 16–17.

Lilley T (2006) Obituary for Professor Anton Linder Hales. Research School of Earth Science, Australian National University.

Lindsay R (1970) *Lord Rayleigh – The Man and His Work*. New York: Pergamon Press.

Manneville P (2006) Rayleigh–Bénard convection: Thirty years of experimental, theoretical, and modeling work. In: Mutabazi I, Wesfried J, and Guyon E (eds.) *Dynamics of Spatio-Temporal Cellular Structure – Henri Bénard Centenary Review*, pp. 41–65. New York: Springer.

McGovern P and Schubert G (1989) Thermal evolution of the earth: Effects of volatile exchange between atmosphere and interior. *Earth and Planetary Science Letters* 96: 27–37.

Menard H (1986) *The Ocean of Truth: A Personal History of Global Tectonics*. Princeton, NJ: Princeton University Press.

Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl ER, and Hung S-H (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303: 338–343.

Nature (1997) News in brief: Kick-boxer jailed for death of geophysicist. *Nature* 389: 657.

Nicolis G (1995) *Introduction to Nonlinear Science*. Cambridge: Cambridge University Press.

Normand C, Pomeau Y, and Velarde MG (1977) Convective instability: A physicist's approach. *Reviews of Modern Physics* 49: 581–624.

O'Connell R, Gable C, and Hager B (1991) Toroidal-poloidal partitioning of lithospheric plate motion. In: Sabadini SR, Lambeck K, and Boschi E (eds.) *Glacial Isostasy, Sea Level and Mantle Rheology*, pp. 535–551. Norwell, MA: Kluwer Academic.

Oxburgh E and Turcotte D (1978) Mechanisms of continental drift. *Reports on Progress in Physics* 41: 1249–1312.

Peacock S (1990) Fluid processes in subduction zones. *Science* 248: 329–337.

Pearson J (1958) On convection cells induced by surface tension. *Journal of Fluid Mechanics* 4: 489–500.

Pekeris C (1935) Thermal convection in the interior of the Earth. *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement* 3: 343–367.

Prout W (1834) *Bridgewater Treatises*. vol. 8, p. 65, London: W. Pickering.

Runcorn S (1962a) Towards a theory of continental drift. *Nature* 193: 311–314.

Runcorn S (1962b) Convection currents in the Earth's mantle. *Nature* 195: 1248–1249.

Runcorn S (1963) Satellite gravity measurements and convection in the mantle. *Nature* 200: 628–630.

Runcorn S (1974) On the forces not moving lithospheric plates. *Tectonophysics* 21: 197–202.

Schubert G, Turcotte D, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.

Strutt, John William (Lord Rayleigh) (1916) On convective currents in a horizontal layer of fluid when the higher temperature is on the under side. *Philosophical Magazine* 32: 529–546.

Strutt, Robert John (Fourth Baron Rayleigh) (1968) *Life of John William Strutt, Third Baron Rayleigh*. Madison, WI: University of Wisconsin Press.

Thompson, Benjamin (Count Rumford) (1797) The propagation of heat in fluids. *Bibliotheque Britannique (Science et Arts)* 90: 97–200, reprinted in Brown, 1967.

Tozer D (1985) Heat transfer and planetary evolution. *Geophysical Surveys* 7: 213–246.

Turcotte D and Oxburgh E (1972) Mantle convection and the new global tectonics. *Annual Review of Fluid Mechanics* 4: 33–66.

Turekian KK and Narendra BL (2002) Earth sciences. In: Altman S (ed.) *Science at Yale*, pp. 157–170. New Haven, CT: Yale University.

van der Hilst R, Widiyantoro S, and Engdahl E (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386: 578–584.

Walker J, Hayes P, and Kasting J (1981) A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86: 9776–9782.

Ward P and Brownlee D (2000) *Rare Earth*. New York: Copernicus/Springer Verlag.

Wegener A (1924) *The Origin of Continents and Oceans*. London: Methuen and co., Ltd.

Wesfried J (2006) Scientific biography of Henri Bénard (1874–1939). In: Mutabazi I, Wesfried J, and Guyon E (eds.) *Dynamics of Spatio-Temporal Cellular Structure – Henri Bénard Centenary Review*, pp. 9–37. New York: Springer.

Williams Q and Hemley R (2001) Hydrogen in the deep Earth. *Annual Review of Earth and Planetary Sciences* 29: 365–418.

Wolfe C, Solomon S, Laske G, et al. (2009) Mantle shear-wave velocity structure beneath the Hawaiian hotspot. *Science* 326: 1388–1390.

7.02 Physics of Mantle Convection

Y Ricard, Ecole Normale Supérieure de Lyon, Université de Lyon-1, CNRS, Lyon, France

© 2015 Elsevier B.V. All rights reserved.

7.02.1	**Introduction**	24
7.02.2	**Conservation Equations**	25
7.02.2.1	General Expression of Conservation Equations	25
7.02.2.2	Mass Conservation	26
7.02.2.3	Momentum Conservation	26
7.02.2.3.1	General momentum conservation	26
7.02.2.3.2	Inertia and non-Galilean forces	27
7.02.2.3.3	Angular momentum conservation	27
7.02.2.4	Energy Conservation	28
7.02.2.4.1	First law and internal energy	28
7.02.2.4.2	State variables	28
7.02.2.4.3	Temperature	29
7.02.2.4.4	Second law and entropy	29
7.02.2.5	Gravitational Forces	30
7.02.2.5.1	Poisson's equation	30
7.02.2.5.2	Self-gravitation	30
7.02.2.5.3	Conservative forms of momentum and energy equations	30
7.02.2.6	Boundary and Interface Conditions	31
7.02.2.6.1	General method	31
7.02.2.6.2	Interface conditions in the 1-D case and for bounded variables	32
7.02.2.6.3	Phase change interfaces and stress continuity?	32
7.02.2.6.4	Weakly deformable surface of a convective cell	33
7.02.3	**Thermodynamic and Rheological Properties**	34
7.02.3.1	Equation of State and Solid Properties	34
7.02.3.2	Rheology	35
7.02.3.2.1	Elasticity	35
7.02.3.2.2	Viscous Newtonian rheology	36
7.02.3.2.3	Maxwellian viscoelasticity	36
7.02.3.2.4	Nonlinear rheologies	37
7.02.4	**Physics of Convection**	38
7.02.4.1	Basic Balance	38
7.02.4.2	Two Simple Solutions	38
7.02.4.2.1	The diffusive solution	38
7.02.4.2.2	The adiabatic solution	39
7.02.4.2.3	Stability of the adiabatic gradient	39
7.02.4.3	Approximate Equations	39
7.02.4.3.1	Depth-dependent reference profiles	39
7.02.4.3.2	Perturbations of the hydrostatic, adiabatic solution	40
7.02.4.3.3	Anelastic approximation	40
7.02.4.3.4	Anelastic liquid approximation	41
7.02.4.3.5	Nondimensionalization	42
7.02.4.3.6	Dimensionless numbers	42
7.02.4.3.7	Boussinesq approximation	42
7.02.4.3.8	Internal heating	43
7.02.4.3.9	Change of nondimensionalization	43
7.02.4.4	Linear Stability Analysis for Basally Heated Convection	43
7.02.4.5	Road to Chaos	44
7.02.5	**Introduction to Physics of Multicomponent and Multiphase Flows**	44
7.02.5.1	Fluid Dynamics of Multicomponent Flows in Solution	45
7.02.5.1.1	Mass conservation in a multicomponent solution	45
7.02.5.1.2	Momentum and energy in a multicomponent solution	46
7.02.5.1.3	Entropy conservation in a multicomponent solution	46
7.02.5.1.4	Advection–diffusion equation and reaction rates	47
7.02.5.1.5	Conservation properties of the advection–diffusion equation	48

7.02.5.1.6 Laminar and turbulent stirring 49
7.02.5.1.7 Diffusion in Lagrangian coordinates 50
7.02.5.2 Fluid Dynamics of Two-Phase Flows 51
7.02.5.2.1 Mass conservation for matrix and fluid 51
7.02.5.2.2 Momentum conservation of matrix and fluid 51
7.02.5.2.3 Energy conservation for two-phase flows 52
7.02.5.2.4 Entropy production and phenomenological laws 53
7.02.5.2.5 Summary equations 54
7.02.6 Specifics of Earth's Mantle Convection 55
7.02.6.1 A Negligible Inertia 55
7.02.6.1.1 Dynamic models 55
7.02.6.1.2 Mantle flow and postglacial models 55
7.02.6.1.3 Time-dependent models 56
7.02.6.2 A Mantle with Internal Heating 56
7.02.6.3 A Complex Rheology 58
7.02.6.3.1 Temperature dependence of viscosity 59
7.02.6.3.2 Depth dependence of viscosity 59
7.02.6.3.3 Stress dependence of viscosity 59
7.02.6.3.4 Grain size dependence of viscosity 60
7.02.6.4 Importance of Sphericity 60
7.02.6.5 Other Depth-Dependent Parameters 60
7.02.6.5.1 Thermal expansivity variations 60
7.02.6.5.2 Increase in average density with depth 60
7.02.6.5.3 Thermal conductivity variations 61
7.02.6.6 Thermochemical Convection 61
7.02.6.6.1 Density variations in the mantle 61
7.02.6.6.2 Phase changes 61
7.02.6.6.3 Abyssal layers 61
7.02.6.7 A Complex Lithosphere: Plates and Continents 62
7.02.6.8 Super-Earths 64
Acknowledgment 64
References 65

7.02.1 Introduction

In many textbooks of fluid dynamics, and for most students, the word 'fluid' refers to one of the states of matter, either liquid or gaseous, in contrast to the solid state. This definition is much too restrictive. In fact, the definition of a fluid rests in its tendency to deform irrecoverably. Basically, any material that appears as elastic or nondeformable, with a crystalline structure (i.e., belonging to the solid state) or with a disordered structure (e.g., a glass, which from a thermodynamic point of view belongs to the liquid state), can be deformed when subjected to stresses for a long enough time.

The characteristic time constant of the geologic processes related to mantle convection, typically 10 My (3×10^{14} s), is so long that the mantle, although stronger than steel and able to transmit seismic shear waves, can be treated as a fluid. Similarly, ice, which is the solid form of water, is able to flow from mountain tops to valleys in the form of glaciers. A formalism that was developed for ordinary liquids or gases can therefore be used in order to study the inside of planets. It is not the equations themselves, but their parameters (viscosity, conductivity, spatial dimensions, etc.) that characterize their applicability to mantle dynamics.

Most materials can therefore behave like elastic solids on very short time constants and like liquids at long times. The characteristic time that controls the appropriate rheological behavior is the ratio between dynamic viscosity, η, and elasticity (shear modulus), μ_r, called the Maxwell time τ_M (Maxwell, 1831–79),

$$\tau_M = \frac{\eta}{\mu_r} \quad [1]$$

The rheological transition in some materials like silicon putty occurs in only a few minutes; a silicon ball can bounce on the floor, but it turns into a puddle when left on a table for tens of minutes. The transition time is of the order of a few hundred to a few thousand years for the mantle (see Section 7.02.3.2). Phenomena of a shorter duration than this time will experience the mantle as an elastic solid, while most tectonic processes will experience the mantle as an irreversibly deformable fluid. Surface loading of the Earth by glaciation and deglaciation involves times of a few thousand years for which elastic aspects cannot be totally neglected with respect to viscous aspects.

Although the word 'convection' is often reserved for flows driven by internal buoyancy anomalies of thermal origin, in this chapter, we will more generally use 'convection' for any motion of a fluid driven by internal or external forcing. Convection can be kinematically forced by boundary conditions or induced by density variations. The former is 'forced convection' and the latter is 'free convection,' which can be of compositional or thermal origin. We will, however, mostly focus

on the aspects of thermal convection (or Rayleigh–Bénard convection (Rayleigh, 1842–1919; Bénard, 1874–1939)) when the fluid motion is driven by thermal anomalies and discuss several common approximations that are made in this case. We know, however, that many aspects of mantle convection can be more complex and involve compositional and petrologic density anomalies or multiphase physics. We will therefore review some of these complexities.

The physics of fluid behavior, like the physics of elastic media, is based on the general continuum hypothesis. This hypothesis requires that quantities like density, temperature, and velocity are defined everywhere, continuously, and at 'points' or infinitesimal volumes that have a statistically meaningful number of molecules so that the quantity represents an average independent of microscopic molecular fluctuations. This hypothesis seems natural for ordinary fluids at the laboratory scale. We will adopt the same hypothesis for the mantle, although we know that it is heterogeneous at various scales and made of compositionally distinct mineral grains.

7.02.2 Conservation Equations

The basic equations of this section can be found in more detail in many classical textbooks (Batchelor, 1967; Landau and Lifchitz, 1980). We will only emphasize the aspects that are pertinent to Earth's and terrestrial mantles.

7.02.2.1 General Expression of Conservation Equations

Let us consider a fluid transported by the velocity field $\boldsymbol{\nu}$, a function of position $\mathbf{x}$ and time t. There are two classical approaches to describe the physics in this deformable medium. Any variable A in a flow (like a scalar such as temperature or a vector like velocity) can be considered as a simple function of position and time, $A(\mathbf{x}, t)$, in a way very similar to the specification of an electromagnetic field. This description is the Eulerian point of view (Euler, 1707–83). The second point of view is traditionally attributed to Lagrange (Lagrange, 1736–1813) and considers that a material element of the flow is identified by its unique initial position $\mathbf{x}_0$ and thereafter follows a trajectory specific to that element $\mathbf{x}(\mathbf{x}_0, t)$. An observer following this trajectory would naturally choose the variable $A(\mathbf{x}(\mathbf{x}_0, t), t)$. The same variable A seen by an Eulerian or a Lagrangian observer would have very different time derivatives. According to Euler, the time derivative would simply be the rate of change seen by an observer at a fixed position, that is, the partial derivative $\partial/\partial t$. According to Lagrange, the time derivative, noted with D, would be the rate of change seen by an observer riding on the material element

$$\frac{\mathrm{D}A}{\mathrm{D}t} = \left(\frac{\mathrm{d}A(\mathbf{x}(\mathbf{x}_0,\ t),t)}{\mathrm{d}t}\right)_{x_0} = \sum_{i=1,3} \frac{\partial A}{\partial x_i}\frac{\partial x_i}{\partial t} + \frac{\partial A}{\partial t} \qquad [2]$$

where x_i are the components of $\mathbf{x}$. Since $\mathbf{x}$ is the position of a material element of the flow, its partial time derivative is simply the flow velocity $\mathbf{v}$. The Lagrangian derivative is also sometimes called material derivative, total derivative, or substantial derivative.

The previous relation was written for a scalar field A, but it could easily be applied to a vector field $\mathbf{A}$. The Eulerian and Lagrangian time derivatives are thus related by the symbolic relation

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + (\mathbf{v}\cdot\boldsymbol{\nabla}) \qquad [3]$$

In Cartesian coordinates (Descartes, 1596–1650), the operator $(\mathbf{v}\cdot\boldsymbol{\nabla})$ is the symbolic scalar $(v_1\partial/\partial x_1 + v_2\partial/\partial x_2 + v_3\partial/\partial x_3)$ and a convenient mnemonic is to interpret it as the scalar product of the velocity field with the gradient operator $(\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$ denoted with the symbol $\boldsymbol{\nabla}$ called nabla. The operator $(\mathbf{v}\cdot\boldsymbol{\nabla})$ can be applied to a scalar or a vector. Notice however the limitations of this mnemonic as for any vectors $\mathbf{A}$ and $\mathbf{B}$, $(\mathbf{v}\cdot\mathbf{B})\mathbf{A}$ is parallel to $\mathbf{A}$, while $(\mathbf{v}\cdot\boldsymbol{\nabla})\mathbf{A}$ is parallel neither to $\mathbf{A}$ nor to $\mathbf{v}$.

In a purely homogeneous fluid, the flow lines are not visible and the mechanical properties are independent of the original position of fluid particles. In this case, the Eulerian perspective seems natural. On the other hand, a physicist describing elastic media can easily draw marks on the surface of deformable objects and flow trajectories become perceptible for him or her. After elastic deformation, the stresses are also dependent on the initial equilibrium state. The Lagrangian perspective is therefore more appropriate. We will mostly adopt the Eulerian perspective for the description of the mantle. However, when we discuss deformation of heterogeneities embedded in and stirred by the convective mantle, the Lagrangian point of view will be more meaningful (see Section 7.02.5.1.7).

A starting point for describing the physics of a continuum is the conservation equations. Consider a scalar- or a vector-extensive variable A (i.e., mass, momentum, energy, entropy, and number of moles) with a density per unit volume ρ_A (with units of A per unit volume) and a virtual but fixed volume Ω enclosed by the open surface $\sum$. This virtual volume is freely crossed by the flow. The temporal change of the net quantity of A inside Ω is

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega \rho_A\ \mathrm{d}V = \int_\Omega \frac{\partial \rho_A}{\partial t}\ \mathrm{d}V \qquad [4]$$

Since Ω is fixed, the derivative of the integral is the integral of the partial time derivative.

The total quantity of the extensive variable A in a volume Ω can be related to a local production, H_A (with units of A per unit volume and unit time), and to the transport (influx or efflux) of A across the interface. This transport can be either a macroscopic advective transport by the flow or a more indirect transport, for example, microscopic diffusion. Let us call $\mathbf{J}_A$ the total flux of A per unit surface area. The conservation of A can be expressed in integral form as

$$\int_\Omega \frac{\partial \rho_A}{\partial t}\ \mathrm{d}V = -\int_\Sigma \mathbf{J}_A\cdot\ \mathrm{d}\mathbf{S} + \int_\Omega H_A\ \mathrm{d}V \qquad [5]$$

where the infinitesimal surface element vector $\mathrm{d}\mathbf{S}$ is oriented with the outward unit normal; hence, the minus sign associates outward flux with a sink of quantity A. Equation [5] is the general form of any conservation equation. When the volume Ω, surface $\sum$, and flux $\mathbf{J}_A$ are regular enough (in mathematical terms when the volume is compact, the surface piecewise

smooth, and the flux continuously differentiable), we can make use of the divergence theorem

$$\int_{\Sigma} \mathbf{J}_A \cdot d\mathbf{S} = \int_{\Omega} \nabla \cdot \mathbf{J}_A \, dV \qquad [6]$$

to transform the surface integral into a volume integral. The divergence of the vector $\mathbf{J}$ with Cartesian coordinates (J_1, J_2, J_3) is the scalar $\nabla \cdot \mathbf{J} = \partial J_1/\partial x_1 + \partial J_2/\partial x_2 + \partial J_3/\partial x_3$, scalar product of the symbolic operator ∇ by the real vector $\mathbf{J}$ (although notice the difference between the scalar $\nabla \cdot \mathbf{J}$ and the operator $\mathbf{J} \cdot \nabla$). Since the integral equation [5] is valid for any virtual volume Ω, we can deduce that the general differential form of the conservation equation is

$$\frac{\partial \rho_A}{\partial t} + \nabla \cdot \mathbf{J}_A = H_A \qquad [7]$$

A similar expression can be used for a vector quantity $\boldsymbol{\rho}_A$ with a tensor flux $\underline{\mathbf{J}}_A$ and a vector source term $\mathbf{H}_A$. In this case, the divergence of the second-order tensor with components J_{ij} is a vector with Cartesian components $\Sigma_{j=1,3} \partial J_{ij}/\partial x_j$ (our convention is that the derivatives in the divergence affect the second index of the tensor; other authors (e.g., Malvern, 1969) use a different convention).

We now apply this formalism to various physical quantities. Three quantities are strictly conserved: the mass, the momentum, and the energy. This means that they can only change in a volume Ω by influx or efflux across the surface $\sum$. We must identify the corresponding fluxes, but no source terms should be present (in fact, in classical mechanics, the radioactivity appears as a source of energy) (see also **Section 7.02.2.5.3**). One very important physical quantity is not conserved – the entropy – but the second law of thermodynamics ensures the positivity of the associated sources.

7.02.2.2 Mass Conservation

Mass flux of material with density ρ (mass per unit volume) is

$$\mathbf{J}_\rho = \rho \mathbf{v} \qquad [8]$$

Using either the Eulerian or the Lagrangian time derivatives, mass conservation becomes

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \qquad [9]$$

or

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0 \qquad [10]$$

In an incompressible fluid, particles have constant density, and so in the particle frame of reference, the Lagrangian observer does not see any density variation and $D\rho/Dt = 0$. In this case, mass conservation takes the simple form $\nabla \cdot \mathrm{v} = 0$, which is commonly called the continuity equation.

Using mass conservation, a few identities can be derived that are very useful for transforming an equation of conservation for a quantity per unit mass to a quantity per unit volume. For example, for any specific scalar quantity A (i.e., quantity per unit mass),

$$\frac{\partial (\rho A)}{\partial t} + \nabla \cdot (\rho A \mathbf{v}) = \rho \frac{DA}{Dt} \qquad [11]$$

and for any specific vector field $\mathbf{A}$,

$$\frac{\partial (\rho \mathbf{A})}{\partial t} + \nabla \cdot (\rho \mathbf{A} \otimes V) = \rho \frac{D\mathbf{A}}{Dt} \qquad [12]$$

where $\mathbf{A} \otimes \mathbf{v}$ is a dyadic tensor of components $A_i v_j$ (notice that $\mathbf{v} \otimes \mathbf{A} \neq \mathbf{A} \otimes \mathbf{v}$).

7.02.2.3 Momentum Conservation

7.02.2.3.1 General momentum conservation

The rate of change of momentum of material in a volume Ω is related to the body forces acting in the volume Ω and to the surface forces acting on its surface $\sum$, that is, Newton's second law (Newton, 1642–1727). The total momentum is

$$\int_{\Omega} \rho \mathbf{v} \, dV \qquad [13]$$

and its variations are due to

- advective transport of momentum across the surface $\sum$,
- forces acting on this surface,
- internal body forces.

The momentum conservation of an open, fixed volume can therefore be expressed in integral form as

$$\int_{\Omega} \frac{\partial (\rho \mathbf{v})}{\partial t} dV = -\int_{\Sigma} \rho \mathbf{v} (\mathbf{v} \cdot d\mathbf{S}) + \int_{\Sigma} \underline{\boldsymbol{\sigma}} \cdot d\mathbf{S} + \int_{\Omega} \mathbf{f} \cdot dV \qquad [14]$$

The stress tensor $\underline{\boldsymbol{\sigma}}$ corresponds to the traction applied on the surface, per unit area (see **Figure 1**). Our convention is that σ_{ij} is the i-component of the force per unit area across a plane normal to the j-direction. The term $\mathbf{f}$ represents the sum of all body forces per unit volume and in particular the gravitational forces $\rho \mathbf{g}$ (we will not consider electromagnetic forces that are described in vol. 8).

Using the divergence theorem (for the first term of the right side, $\rho \mathbf{v} (\mathbf{v} \cdot d\mathbf{S})$ can also be written as $\rho (\mathbf{v} \otimes \mathbf{v}) \cdot d\mathbf{S}$) and the equality [12], the differential form of momentum conservation becomes

$$\rho \frac{D\mathbf{v}}{Dt} = \nabla \cdot \underline{\boldsymbol{\sigma}} + \mathbf{f} \qquad [15]$$

It is common to divide the total stress tensor into a thermodynamic pressure – $P\underline{\mathbf{I}}$ where $\underline{\mathbf{I}}$ is the identity stress tensor and a

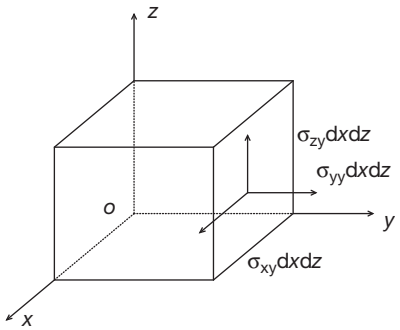


Figure 1 The force per unit area applied on a surface directed by the normal vector $\mathbf{n}$ is by definition $\underline{\boldsymbol{\sigma}} \cdot \mathbf{n}$. The component of this force along the unit vector $\mathbf{e}_i$, therefore $\mathbf{e}_i \cdot \underline{\boldsymbol{\sigma}} \cdot \mathbf{n}$. For example, $\sigma_{zy} dxdz$ is the component acting in the oz direction of force applied on the surface $dxdz$ with normal in the oy direction. Here again, the order of the indexes for the stress tensor is a convention but in agreement with our definition of the divergence of a tensor.

velocity-dependent stress $\underline{\boldsymbol{\tau}}$. The relationship between the tensor $\underline{\boldsymbol{\tau}}$ and the velocity field will be discussed in Section 7.02.3.2. Without motion, the total stress tensor is thus isotropic and equal to the usual pressure. In most geophysical literature, it has been assumed that the velocity-dependent tensor has no isotropic component, that is, it is traceless $\mathrm{tr}(\underline{\boldsymbol{\tau}}) = 0$. In this case, the average isotropic total stress $-\mathrm{tr}(\underline{\boldsymbol{\sigma}})/3$ is the thermodynamic pressure P and the velocity-dependent stress tensor $\underline{\boldsymbol{\tau}}$ is the deviatoric stress tensor (see Section 7.02.3.2 for more details).

As $\boldsymbol{\nabla}\cdot(P\underline{\mathbf{I}}) = \boldsymbol{\nabla}P$, momentum conservation [14], in terms of pressure and deviatoric stresses, is

$$\rho\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = -\boldsymbol{\nabla}P + \boldsymbol{\nabla}\cdot\underline{\boldsymbol{\tau}} + \mathbf{f} \qquad [16]$$

This equation is called the Navier–Stokes equation (Navier, 1785–1836; Stokes, 1819–1903) when the stress tensor is linearly related to the strain rate tensor and the fluid is incompressible (see Section 7.02.3.2).

7.02.2.3.2 Inertia and non-Galilean forces

In almost all studies of mantle dynamics, the fact that the Earth is rotating is simply neglected. It is however worth discussing this point. Let us define a reference frame of vectors $\mathbf{e}_i$ attached to the solid Earth. These vectors rotate with the Earth, such that with respect to a Galilean frame, their rate of change is

$$\frac{\mathrm{d}\mathbf{e}_i}{\mathrm{d}t} = \boldsymbol{\omega}\times\mathbf{e}_i \qquad [17]$$

where $\boldsymbol{\omega}$ is the angular velocity of Earth's rotation. A point of the Earth, $\mathbf{x} = \sum_i x_i\mathbf{e}_i$, has a velocity in the Galilean frame

$$\left[\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\right]_{\mathrm{Gal.}} = \sum_{i=1,3}\left[\left(\frac{\mathrm{d}x_i}{\mathrm{d}t}\right)\mathbf{e}_i + x_i\left(\frac{\mathrm{d}\mathbf{e}_i}{\mathrm{d}t}\right)\right] = \mathbf{v}_{\mathrm{Earth}} + \boldsymbol{\omega}\times\mathbf{x} \qquad [18]$$

where $\mathbf{v}_{\mathrm{Earth}}$ is the velocity in the Earth's frame, and, by repeating the Galilean time derivate, to obtain the acceleration $\boldsymbol{\gamma}_{\mathrm{Gal}}$ in a Galilean frame (Galileo, 1564–1642),

$$\boldsymbol{\gamma}_{\mathrm{Gal}} = \boldsymbol{\gamma}_{\mathrm{Earth}} + 2\boldsymbol{\omega}\times\mathbf{v}_{\mathrm{Earth}} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\mathbf{x}) + \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t}\times\mathbf{x} \qquad [19]$$

In this well-known expression, one recognizes on the right side the acceleration in the non-Galilean Earth's reference frame and the Coriolis (Coriolis, 1792–1843), centrifugal, and Poincaré accelerations (Poincaré, 1854–1912).

To quantify the importance of the first three acceleration terms (neglecting the Poincaré term), let us consider a characteristic length scale (the Earth's radius, $a = 6371$ km) and mantle velocity (the maximum plate tectonic speed, $U = 10$ cm year^{-1}) with which we can compare the various acceleration terms. One immediately gets

$$\frac{\text{Inertia}}{\text{Coriolis}} = \frac{U}{2\omega a} = \frac{1}{2.9\times10^{11}} \qquad [20]$$

$$\frac{\text{Coriolis}}{\text{Gravitational force}} = \frac{2\omega U}{g} = \frac{1}{2.1\times10^{13}} \qquad [21]$$

$$\frac{\text{Centrifugal}}{\text{Gravitational force}} = \frac{\omega^2 a}{g} = \frac{1}{291} \qquad [22]$$

Thus, the inertial term is much smaller than the Coriolis term (this ratio is also known as the Rossby number (Rossby, 1898–1957)), which is itself negligible relative to gravitational force. Even if we argue that a more meaningful comparison would be between the whole Coriolis force $2\rho\omega U$ and the lateral variations of the gravitational force $\delta\rho g$ rather than the total gravity force ρg, inertia and Coriolis accelerations still play a negligible role in mantle dynamics (assuming that the mantle velocity U is related to $\delta\rho g$ by a viscous relation $U \propto \delta\rho g a^2/\eta$ where η is the mantle viscosity, the ratio $2\rho\omega U/(\delta\rho g)$ is the inverse of the Ekman number (Ekman, 1874–1954)). Neglecting inertia means that forces are instantaneously in balance and that changes in kinetic energy are negligible since inertia is the time derivative of the kinetic energy. We can perform a simple numerical estimate of the mantle kinetic energy. The kinetic energy of a lithospheric plate (a square of size 2000 km, thickness 100 km, velocity 5 cm year^{-1}, and density 3000 kg m^{-3}) is 1.51 kJ, which is comparable to that of a medium-sized car (2000 kg) driven at only 4.42 km h^{-1} or of a man on a bike (say, total about 100 kg) going at a leisurely 20 km h^{-1}.

The centrifugal term is also quite small but not as small (i.e., it is 1/291 of gravitational force). It induces several effects. The centrifugal effect causes the Earth's flattening with an equatorial bulge of 21 km (1/300 of the Earth's radius), although this is a static phenomenon that has no interactions with convective dynamics. The gravitational coupling between the equatorial bulge and other celestial bodies perturbs their Keplerian orbits (Kepler, 1571–1630) and leads to precessions and nutations of their orbits as well as Earth's spin. Finally, the bulge causes the whole planet to rotate along an equatorial axis in order to keep the maximum principal inertial axis coincident with its rotational axis (Ricard et al., 1993b; Spada et al., 1992a). This rotational equilibrium of the Earth will not be discussed here (see e.g., Chandrasekhar (1969) for the static equilibrium shape of a rotating planet and Munk and MacDonald (1960) for the dynamics of a deformable rotating body). The Poincaré acceleration is itself negligible compared to the centrifugal term as the changes of Earth's rotation occur over times much larger than a day. Although Poincaré acceleration has no effect on the mantle, it may affect the dynamics of the fluid core (Vanyo et al., 1995).

We neglect all the acceleration terms in the following, but we should remember that in addition to the convective motion of a nonrotating planet, a rotation of the planet with respect to an equatorial axis is possible. This motion documented by paleomagnetism is called true polar wander (Besse and Courtillot, 1991).

7.02.2.3.3 Angular momentum conservation

The angular momentum per unit mass $\mathbf{J} = \mathbf{x}\times\mathbf{v}$ also obeys a conservation law. This law can be obtained in two different ways. First, as we did for mass and momentum conservation, we can express the balance of angular momentum in integral form. In the absence of intrinsic body couples (i.e., we do not consider the case where distributed body couples exist in addition of distributed body forces), the angular momentum variations are due to

- advective transport of angular momentum across the surface $\sum$,
- torque of forces acting on this surface,
- torque of internal body forces.

The resulting balance is therefore

$$\int_\Omega \frac{\partial(\rho\mathbf{J})}{\partial t}\,\mathrm{d}V = -\int_\Sigma \rho\mathbf{J}(\mathbf{v}\cdot\mathrm{d}\mathbf{S}) + \int_\Sigma \mathbf{x}\times(\underline{\boldsymbol{\sigma}}\cdot\mathrm{d}\mathbf{S}) + \int_\Omega \mathbf{x}\times\mathbf{f}\,\mathrm{d}V \quad [23]$$

The only difficulty to transform this integral form into a local equation is with the integral involving the stress tensor (but is readily solved using tensor or index notation; see Malvern, 1969). After some algebra, eqn [23] becomes

$$\rho\frac{\mathrm{D}\mathbf{J}}{\mathrm{D}t} = \mathbf{x}\times\nabla\cdot\underline{\boldsymbol{\sigma}} + \mathbf{x}\times\mathbf{f} + \mathcal{T} \quad [24]$$

where the torque $\mathcal{T}$ is the vector $(\mathcal{T}_{zy}-\mathcal{T}_{yz}, \mathcal{T}_{xz}-\mathcal{T}_{zx}, \mathcal{T}_{yx}-\mathcal{T}_{xy})$. A second expression can be obtained by the vectorial multiplication of the momentum equation [15] by $\mathbf{x}$. Since

$$\mathbf{x}\times\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = \frac{\mathrm{D}\mathbf{J}}{\mathrm{D}t} - \frac{\partial\mathbf{x}}{\partial t}\times\mathbf{v} = \frac{\mathrm{D}\mathbf{J}}{\mathrm{D}t} - \mathbf{v}\times\mathbf{v} = \frac{\mathrm{D}\mathbf{J}}{\mathrm{D}t} \quad [25]$$

we get,

$$\rho\frac{\mathrm{D}\mathbf{J}}{\mathrm{D}t} = \mathbf{x}\times\nabla\cdot\underline{\boldsymbol{\sigma}} + \mathbf{x}\times\mathbf{f} \quad [26]$$

which differs from eqn [24] by the absence of the torque $\mathcal{T}$. This proves that in the absence of singular sources of angular momentum (distributed body couples), the stress (either $\underline{\boldsymbol{\sigma}}$ or $\underline{\boldsymbol{\tau}}$) must be represented by a symmetrical tensor,

$$\underline{\boldsymbol{\sigma}} = \underline{\boldsymbol{\sigma}}^t, \quad \underline{\boldsymbol{\tau}} = \underline{\boldsymbol{\tau}}^t \quad [27]$$

where $[\,]^t$ denotes tensor transposition.

7.02.2.4 Energy Conservation

7.02.2.4.1 First law and internal energy

The total energy per unit mass of a fluid is the sum of its internal energy, $\mathcal{U}$, and its kinetic energy (this approach implies that the work of the various forces is separately taken into account; another approach that we use in Section 7.02.2.5.3 adds to the total energy the various possible potential energies and ignores forces). In the fixed volume Ω, the total energy is thus

$$\int_\Omega \rho\left(\mathcal{U} + \frac{v^2}{2}\right)\mathrm{d}V \quad [28]$$

A change of this energy content can be caused by

- advection of energy across the boundary $\sum$ by macroscopic flow;
- transfer of energy through the same surface without mass transport, by say diffusion or conduction;
- work of body forces;
- work of surface forces;
- volumetrically distributed radioactive heat production.

Using the divergence theorem, the balance of energy can therefore be written as

$$\frac{\partial}{\partial t}\left(\rho\left(\mathcal{U}+\frac{v^2}{2}\right)\right) = -\nabla\cdot\left(\rho\left(\mathcal{U}+\frac{v^2}{2}\right)\mathbf{v} + \mathbf{q} + P\mathbf{v} - \underline{\boldsymbol{\tau}}\cdot\mathbf{v}\right) + \mathbf{f}\cdot\mathbf{v} + \rho H \quad [29]$$

where $\mathbf{q}$ is the diffusive flux, H is the rate of energy production per unit mass, and the stresses are divided into thermodynamic pressure- and velocity-dependent stresses.

This expression can be developed and simplified by using eqn [11] and the equations of mass and momentum conservation [9] and [16] to reach the form

$$\rho\frac{\mathrm{D}\mathcal{U}}{\mathrm{D}t} = -\nabla\cdot\mathbf{q} - P\nabla\cdot\mathbf{v} + \underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \rho H \quad [30]$$

The viscous dissipation term $\underline{\boldsymbol{\tau}}:\nabla\mathbf{v}$ is the contraction of the two tensors $\underline{\boldsymbol{\tau}}$ and $\nabla\mathbf{v}$ (of components $\partial v_i/\partial x_j$). Its expression is $\Sigma_{ij}\tau_{ij}\partial v_i/\partial x_j$.

7.02.2.4.2 State variables

The internal energy can be expressed in terms of the more usual thermodynamic state variables, namely, temperature, pressure, and volume. We use volume to follow the classical thermodynamic approach, but since we apply thermodynamics to points in a continuous medium, the volume V is in fact the volume per unit mass or $1/\rho$. We use the first law of thermodynamics that states that during an infinitesimal process, the variation of internal energy is the sum of the heat δQ and work δW exchanged. Although irreversible processes occur in the fluid, we assume that we can adopt the hypothesis of local thermodynamic equilibrium.

The increments in heat and work are not exact differentials: these increments depend not only on the initial and final stages but also on the entire process of energy exchange. Using either a T–V or a T–P formulation, we can write

$$\delta Q = C_V\mathrm{d}T + l\mathrm{d}V = C_p\mathrm{d}T + h\mathrm{d}P \quad [31]$$

where C_P an C_V are the heat capacities per unit mass at constant pressure and volume, respectively, and h and l are two other calorimetric coefficients necessary to account for heat exchange at constant temperature. For fluids, the reversible exchange of work is only due to the work of pressure forces:

$$\delta W = -P\mathrm{d}V \quad [32]$$

which implies that only the pressure term corresponds to an energy that can be stored and recovered without loss when the change in volume is reversed. In contrast, the stresses related to the velocity will ultimately appear in the dissipative, irrecoverable term of viscous dissipation: this point will be further considered in the Section 7.02.3.2 about rheology.

Thermodynamics states that the total variations of energy, $\mathrm{d}\mathcal{U} = \delta Q + \delta W$; enthalpy, $\mathrm{d}\mathcal{H} = \mathrm{d}\mathcal{U} + \mathrm{d}(PV)$; or entropy, $\mathrm{d}\mathcal{S} = \delta Q/T$, are exact differentials and $\mathcal{U}, \mathcal{H}$, and $\mathcal{S}$ are potentials. This means that the net change in energy (or enthalpy or entropy) between an initial state and a final state depends only on the initial and final states themselves and not on the intermediate stages. This implies mathematically that the second partial derivatives of these potentials with respect to any pair of variables are independent of the order of differentiation. Using these rules, a large number of relations can be derived among the thermodynamic coefficients and their derivatives. These are called the Maxwell relations and are discussed in most thermodynamics textbooks (e.g., Poirier, 1991). We can in particular derive the values of l (starting from $\mathrm{d}\mathcal{U}$ and $\mathrm{d}\mathcal{S}$ in the T–V formulation) and h (starting from $\mathrm{d}\mathcal{H}$ and $\mathrm{d}\mathcal{S}$ in the T–P formulation),

$$l = \alpha T K_T, \quad \text{and} \quad h = -\frac{\alpha T}{\rho} \quad [33]$$

In these expressions for l and h, we introduced the thermal expansivity α and the isothermal incompressibility K_T,

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P = -\frac{1}{\rho}\left(\frac{\partial \rho}{\partial T}\right)_P$$
$$K_T = -V\left(\frac{\partial P}{\partial V}\right)_T = \rho\left(\frac{\partial P}{\partial \rho}\right)_T \quad [34]$$

The thermodynamic laws and differentials apply to a closed deformable volume $\Omega(t)$, which corresponds to the perspective of Lagrange. We can therefore interpret the differential symbols 'd' of the thermodynamic definitions [31] and [32] as Lagrangian derivatives 'D.'

Therefore, in total, when the expressions for l and h are taken into account, [33], and when the differential symbols are interpreted as Lagrangian derivatives, the change of internal energy, $d\mathcal{U} = \delta Q + \delta W$, can be recast as

$$\frac{D\mathcal{U}}{Dt} = C_V\frac{DT}{Dt} + (\alpha T K_T - P)\frac{\nabla\cdot\mathbf{v}}{\rho} \quad [35]$$

or

$$\frac{D\mathcal{U}}{Dt} = C_P\frac{DT}{Dt} - \frac{\alpha T}{\rho}\frac{DP}{Dt} - P\frac{\nabla\cdot\mathbf{v}}{\rho} \quad [36]$$

In these equations, we also have replaced the volume variation using mass conservation [9]

$$\frac{DV}{Dt} = \frac{D(1/\rho)}{Dt} = -\frac{1}{\rho^2}\frac{D\rho}{Dt} = \frac{\nabla\cdot\mathbf{v}}{\rho} \quad [37]$$

7.02.2.4.3 Temperature

We can now employ either thermodynamic relation [35] or [36] in our conservation equation deduced from fluid mechanics, [30], to express the conservation of energy in terms of temperature variations:

$$\rho C_P\frac{DT}{Dt} = -\nabla\cdot\mathrm{q} + \alpha T\frac{DP}{Dt} + \underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \rho H$$

$$\rho C_V\frac{DT}{Dt} = -\nabla\cdot\mathrm{q} - \alpha T K_T\nabla\cdot\mathbf{v} + \underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \rho H \quad [38]$$

Apart from diffusion, three sources of temperature variations appear on the right side of these equations. The last term ρH is the source of radioactive heat production. This term is of prime importance for the mantle, mostly heated by the decay of radioactive elements like uranium (^{235}U and ^{238}U), thorium (^{236}Th), and potassium (^{40}K). All together, these nuclides generate about 20×10^{12} W (McDonough and Sun, 1995). Although this number may seem large, it is in fact very small. Since the Earth now has about 7×10^9 inhabitants, the total natural radioactivity of the Earth is only $\approx$3 kW/person, not enough to run the appliances of a standard kitchen in a developed country. It is amazing that this ridiculously small energy source drives plate tectonics, raises mountains, and produces a magnetic field. In addition to the present-day radioactivity, extinct radionuclides, like that of ^{36}Al (with a half-life of 0.73 My), have played an important role in the initial stage of planet formation (Lee et al., 1976).

The viscous dissipation term $\underline{\boldsymbol{\tau}}:\nabla\mathbf{v}$ converts mechanical energy into heat. This term explains the classical Joule experiment (equivalence between work and heat, Joule, 1818–89) in which the potential energy of a load (measured in joules) drives a propeller in a fluid and dissipates the mechanical energy as thermal energy (measured in calories).

The remaining source term, containing the thermodynamic coefficients (α or αK_T in eqn [38]), vanishes when the fluid is incompressible (e.g., when $\alpha = 0$ or when $\nabla\cdot\mathbf{v} = 0$). This term is related to adiabatic compression and will be discussed in Section 7.02.4.2.2.

7.02.2.4.4 Second law and entropy

We now consider the second law of thermodynamics and entropy conservation. Assuming local thermodynamic equilibrium, we have $d\mathcal{U} = T d\mathcal{S} - P dV$. Using the equation of conservation for the internal energy $\mathcal{U}$, [30], and expressing the volume change in terms of velocity divergence, [37], we obtain

$$\rho T\frac{D\mathcal{S}}{Dt} = -\nabla\cdot\mathbf{q} + \underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \rho H \quad [39]$$

To identify the entropy sources, we can express this equation in the form of a conservation equation (see eqn [7]),

$$\frac{\partial(\rho\mathcal{S})}{\partial t} = -\nabla\cdot\left(\rho\mathcal{S}\mathbf{v} + \frac{\mathbf{q}}{T}\right) - \frac{1}{T^2}\mathbf{q}\cdot\nabla T + \frac{1}{T}\underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \frac{1}{T}\rho H \quad [40]$$

The physical meaning of this equation is therefore that the change of entropy is related to a flux of advected and diffused entropy, $\rho\mathcal{S}\mathbf{v}$ and $\mathbf{q}/T$, and to three entropy production terms, including from radiogenic heating.

A brief introduction to the general principles of non-equilibrium thermodynamics will be given in Section 7.02.5.1.4. Here, we simply state that the second law requires that in all situations, the total entropy production is positive. When different entropy production terms involve factors of different tensor orders (tensors, vectors, or scalars), they must separately be positive. This is called the Curie principle (Curie, 1859–1906) (see e.g., de Groot and Mazur, 1984; Woods, 1975). It implies that

$$-\mathbf{q}\cdot\nabla T \geq 0 \quad \text{and} \quad \underline{\boldsymbol{\tau}}:\nabla\mathbf{v} \geq 0 \quad [41]$$

The usual Fourier law (Fourier, 1768–1830) with a positive thermal conductivity $k > 0$,

$$\mathbf{q} = -k\nabla T \quad [42]$$

satisfies the second law.

When the conductivity k is uniform, the thermal diffusion term of the energy equation $-\nabla\cdot\mathbf{q}$ becomes $k\nabla^2 T$ where $\nabla^2 = \nabla\cdot\nabla$ is the scalar Laplacian operator (Laplace, 1749–1827). Instead of a thermal conductivity, a thermal diffusivity κ can be introduced:

$$\kappa = \frac{k}{\rho C_P} \quad [43]$$

(although in principle, isobaric and isochoric thermal diffusivities should be defined). In situations with uniform conductivity, without motion and radioactivity sources, the energy equation [38] becomes the standard diffusion equation

$$\frac{\partial T}{\partial t} = \kappa\nabla^2 T \quad [44]$$

The relation between stress and velocity will be discussed in detail in Section 7.02.3.2. We will show that the relationship

$$\underline{\boldsymbol{\tau}} = 2\eta\left(\underline{\dot{\boldsymbol{\varepsilon}}} - \frac{1}{3}\nabla\cdot\mathbf{v}\underline{\mathbf{I}}\right) \quad [45]$$

is appropriate for the mantle, where η is dynamic viscosity and strain rate tensor $\underline{\dot{\boldsymbol{\varepsilon}}}$ is defined by

$$\underline{\dot{\boldsymbol{\varepsilon}}} = \frac{1}{2}\left([\nabla \mathbf{v}] + [\nabla \mathbf{v}]^t\right) \quad [46]$$

The positivity of the entropy source [41] implies that

$$\underline{\boldsymbol{\tau}} : \underline{\dot{\boldsymbol{\varepsilon}}} = \eta \underline{\dot{\boldsymbol{\varepsilon}}} : \underline{\dot{\boldsymbol{\varepsilon}}} \geq 0 \quad [47]$$

and the dynamic viscosity is therefore a positive quantity. Using this relation and assuming η uniform, the divergence of the stress tensor that appears in the momentum conservation equation has the simple form

$$\nabla \cdot \underline{\boldsymbol{\tau}} = \eta \nabla^2 \mathbf{v} + \frac{\eta}{3} \nabla (\nabla \cdot \mathbf{v}) \quad [48]$$

where the vectorial Laplacian $\nabla^2 \mathbf{v}$ is $\nabla(\nabla \cdot \mathbf{v}) - \nabla \times (\nabla \times \mathbf{v})$. This relationship suggests a meaningful interpretation of the viscosity. The momentum equation [16] can be written as

$$\frac{\partial \mathbf{v}}{\partial t} = \frac{\eta}{\rho} \nabla^2 \mathbf{v} + \text{other terms} \ldots \quad [49]$$

forgetting the other terms, a comparison with the thermal diffusion equation [44] shows that the kinematic viscosity ν, defined by

$$\nu = \frac{\eta}{\rho} \quad [50]$$

should rather be called the momentum diffusivity; it plays the same role with respect to the velocity as thermal diffusivity does with respect to temperature.

7.02.2.5 Gravitational Forces

7.02.2.5.1 Poisson's equation

In this chapter, the only force is the gravitational body force. The Earth's gravity is the sum of this gravitational body force and the centrifugal force already discussed in Section 7.02.2.3.2. The gravitational force per unit mass is the gradient of the gravitational potential ψ, a solution of Poisson's equation (Poisson, 1781–1840), that is,

$$\mathbf{g} = -\nabla \psi, \quad \text{and} \quad \nabla^2 \psi = 4\pi G \rho \quad [51]$$

where G is the gravitational constant. In the force term that appears in the momentum equation [16], $\mathbf{f} = \rho \mathbf{g}$, the gravitational force per unit mass should be in agreement with the distribution of masses: the Earth should be self-gravitating.

7.02.2.5.2 Self-gravitation

When dealing with fluid dynamics at the laboratory scale, the gravitational force can be considered as constant and uniform. The gravitational force is related to the entire distribution of mass in the Earth (and the universe) and is practically independent of the local changes in density in the experimental environment. Therefore, at the laboratory scale, it is reasonable to ignore Poisson's equation and to assume that $\mathbf{g}$ is a uniform and constant reference gravitational field.

Inside a planet, the density can be divided into an average depth-dependent density, $\rho_0(r)$; the source of the reference depth-dependent gravitational field, $\mathbf{g}_0(r)$; and a density perturbation $\delta\rho$, the source of a gravitational perturbation $\delta\mathbf{g}$. The force term $\mathbf{f}$ is therefore to first order $\rho_0 \mathbf{g}_0 + \delta\rho \mathbf{g}_0 + \rho_0 \delta\mathbf{g}$; it is tempting to assume that each term in this expression is much larger than the next one and hopefully that only the first two terms are of importance (neglecting the second term would suppress any feedback between density perturbations and flow). Practically, this assumption would imply consideration of the total density anomalies but only the depth-dependent gravitational field. Solving Poisson's equation to compute the perturbed gravitational field would thus be avoided. We can test the previously mentioned idea and show that unfortunately, the third term, $\rho_0 \delta\mathbf{g}$, may be of the same order as the second one (Panasyuk et al., 1996; Ricard et al., 1984; Richards and Hager, 1984). To perform this exercise, we have to introduce the spherical harmonic functions $Y_{lm}(\theta, \phi)$. These functions of latitude θ and longitude ϕ oscillate on a sphere of radius a just like 2-D sinusoidal functions on a plane. Each harmonic function changes sign $l-m$ times from North to South Pole and m times over the same angle (180°) around the equator. The degree l can thus be interpreted as corresponding to a wavelength of order $2\pi a/l$. Spherical harmonics constitute a basis for functions defined on the sphere and are also eigenfunctions of the angular part of Laplace's equation that facilitates the solution of Poisson's equation.

Let us consider a density anomaly $\delta_\rho = \sigma\delta(r-a)Y_{lm}(\theta, \phi)$ at the surface of a sphere of radius a and uniform density ρ_0, where $\delta(r-a)$ is the Dirac delta function (Dirac (1902–84); σ has unit of kg m^{-2}). Inside the planet, this mass distribution generates the radial gravitational perturbation field of

$$\delta g = 4\pi G \sigma \frac{l}{2l+1} \left(\frac{r}{a}\right)^{l-1} Y_{lm}(\theta, \phi) \quad [52]$$

We can compare the terms $\rho_0 \langle \delta g \rangle$ and $g_0 \langle \delta\rho \rangle$, both averaged over the planet radius. For a uniform planet, the surface gravitational force per unit mass is $g_0 = 4\pi G \rho_0 a/3$. Since $\langle \delta_\rho \rangle = \sigma Y_{lm}(\theta, \phi)/a$, we get

$$\frac{\rho_0 \langle \delta g \rangle}{g_0 \langle \delta\rho \rangle} = \frac{3}{2l+1} \quad [53]$$

This estimate is certainly crude and a precise computation taking into account a distributed density distribution could be done. However, this rule of thumb would remain valid. At low degree, the effect of self-gravitation $\rho_0 \delta\mathbf{g}$ is about 50% of the direct effect $\delta\rho \mathbf{g}_0$ and reaches 10% of it only near $l \approx 15$. Self-gravitation has been taken into account in various models intended to explain Earth's gravity field from mantle density anomalies (see also Ribe, this volume, and Forte, vol. 3). Some spherical convection codes account for self-gravitation (e.g., Zhong et al., 2008); the codes that do not include this effect lead to an inaccurate computation of the gravity field.

7.02.2.5.3 Conservative forms of momentum and energy equations

In the general remarks on conservation laws in Section 7.02.2.1, we wrote that conserved quantities like mass, momentum, and energy can only be transported but do not have production terms (contrary to entropy). However, in the momentum conservation [16] and in the energy conservation [29], two terms, $\mathbf{f} = \rho \mathbf{g}$ and $\mathbf{fv} = \rho \mathbf{g} \cdot \mathbf{v}$, appear as sources (we also said that the radioactive term ρH appears because classical physics does not identify mass as energy and vice versa. A negligible mass sink of order $-\rho H/c^2$, where c is the speed of light, should, moreover, be present in the mass conservation).

It is interesting to check that our equations can be recast into an exact conservative form. An advantage of writing equations in conservative form is that it is appropriate to treat with global balances, interfaces, and boundaries (see Section 7.02.2.6). We can obtain conservative equations by using Poisson's relation [51] and performing some algebra (using $\rho=(\nabla^2\psi)/(4\pi G)$ and $\mathbf{g}\cdot\nabla\mathbf{g}=\nabla\mathbf{g}\cdot\mathbf{g}$, since $\mathbf{g}=-\nabla\psi$):

$$\rho\mathbf{g}=-\frac{\nabla\cdot\mathbf{g}}{4\pi G}\mathbf{g}=-\frac{1}{4\pi G}\nabla\cdot\left(\mathbf{g}\otimes\mathbf{g}-\frac{1}{2}g^2\underline{\mathbf{I}}\right) \quad [54]$$

$$\begin{aligned}\rho\mathbf{g}\cdot\mathbf{v}&=-\rho\mathbf{v}\cdot\nabla\psi\\&=-\rho\left(\frac{\mathrm{D}\psi}{\mathrm{D}t}-\frac{\partial\psi}{\partial t}\right)\\&=-\rho\frac{\mathrm{D}\psi}{\mathrm{D}t}-\frac{\nabla\cdot\mathbf{g}}{4\pi G}\frac{\partial\psi}{\partial t}\\&=-\rho\frac{\mathrm{D}\psi}{\mathrm{D}t}-\frac{1}{4\pi G}\nabla\cdot\left(\mathbf{g}\frac{\partial\psi}{\partial t}\right)-\frac{1}{8\pi G}\frac{\partial g^2}{\partial t}\end{aligned} \quad [55]$$

If we substitute these two expressions in the momentum and the energy conservation equations, [16] and [29], we obtain the conservative forms

$$\frac{\partial(\rho\mathbf{v})}{\partial t}=-\nabla\cdot\left(\rho\mathbf{v}\otimes\mathbf{v}+P\underline{\mathbf{I}}-\underline{\boldsymbol{\tau}}+\frac{1}{4\pi G}\mathbf{g}\otimes\mathbf{g}-\frac{1}{8\pi G}g^2\underline{\mathbf{I}}\right) \quad [56]$$

$$\begin{aligned}&\frac{\partial}{\partial t}\left(\rho\left(\mathcal{U}+\psi+\frac{v^2}{2}\right)+\frac{g^2}{8\pi G}\right)\\&=-\nabla\cdot\left(\rho\mathbf{v}\left(\mathcal{U}+\psi+\frac{v^2}{2}\right)+\mathbf{q}+P\mathbf{v}-\underline{\boldsymbol{\tau}}\cdot\mathbf{v}+\frac{\mathbf{g}}{4\pi G}\frac{\partial\psi}{\partial t}\right)+\rho H\end{aligned} \quad [57]$$

When the gravitational force is time-independent, a potential ψ can simply be added to the kinetic and internal energies to replace the work of gravitational forces. When gravitational force and its potential are time-dependent (due to mass redistribution during convection, segregation of elements, etc.), two new terms must be added, a gravitational energy proportional to g^2 and a gravitational flux proportional to $\mathbf{g}\partial\psi/\partial t$ (this is equivalent to the magnetic energy proportional to B^2, where $\mathbf{B}$ is the magnetic induction, in tesla (Tesla, 1856–1943), and to the Poynting vector of magnetohydrodynamics (Poynting, 1852–1914)).

In a permanent or in a statistically steady regime, the time-dependent terms of energy equation [57] can be neglected and the equation can then be integrated over the volume of the Earth. The natural assumption is that the Earth's surface velocities are perpendicular to the Earth's surface normal vector and that the surface is either stress-free or with no horizontal velocity (we exclude the case of convection forced by imposing a nonzero surface velocity). Using the divergence theorem to transform the volume integral of the divergence back to a surface integral of flux, most terms in eqn [57] cancel and all that remains is

$$\int_\Sigma\mathbf{q}\cdot\mathrm{d}\mathbf{S}=\int_\Omega\rho H\,\mathrm{d}V \quad [58]$$

The surface flux in a statistically steady regime is simply the total radiogenic heat production.

It may seem surprising at first that viscous dissipation that is everywhere positive does not appear in this balance. To understand this point, we can directly integrate the energy equation written in terms of temperature [38],

$$\int_\Omega\rho C_P\frac{\mathrm{D}T}{\mathrm{D}t}\mathrm{d}V=-\int_\Sigma\mathbf{q}\cdot\mathrm{d}\mathbf{S}+\int_\Omega\left(\alpha T\frac{\mathrm{D}P}{\mathrm{D}t}+\underline{\boldsymbol{\tau}}:\nabla\mathbf{v}\right)\mathrm{d}V+\int_\Omega\rho H\mathrm{d}V \quad [59]$$

On the right side, the first and last terms cancel each other out by eqn [58]. To simplify the left side, notice that the enthalpy $\mathcal{H}=\mathcal{U}+P/\rho$ obeys (see eqn [36])

$$\rho\frac{\mathrm{D}\mathcal{H}}{\mathrm{D}t}=\rho C_P\frac{\mathrm{D}T}{\mathrm{D}t}+(1-\alpha T)\frac{\mathrm{D}P}{\mathrm{D}t} \quad [60]$$

According to eqn [11], $\rho\mathrm{D}\mathcal{H}/\mathrm{D}t=\partial(\rho\mathcal{H})/\partial t+\nabla(\rho\mathbf{v}\mathcal{H})$. The last term cancels when integrated on Earth's volume (using the divergence theorem), and the first term cancels in a statistically steady regime. Therefore, the time-average volume integral of $\rho\mathrm{D}\mathcal{H}/\mathrm{D}t$ is zero and the volume integral of $\rho C_p\mathrm{D}T/\mathrm{D}t$ must be opposite to that of $(1-\alpha T)\mathrm{D}P/\mathrm{D}t$. The time average of the energy equation [59] implies therefore (Alboussiere and Ricard, 2013)

$$\int_\Omega\left(\frac{\mathrm{D}P}{\mathrm{D}t}+\underline{\boldsymbol{\tau}}:\nabla\mathbf{v}\right)\mathrm{d}V=0 \quad [61]$$

The total heat production due to dissipation is balanced by the work due to compression and expansion over the convective cycle (notice that if we had assumed a constant C_P, the left side of eqn [59] would have been zero and we would have found an αT term in front of $\mathrm{D}P/\mathrm{D}t$ (Hewitt et al., 1975), which is incorrect). This balance is global not local. Dissipation occurs mostly near the boundary layers of the convection and compressional work is done along the downwellings and upwellings of the flow. The relation [61] can also be derived by multiplying the Navier–Stoke equation [16] by $\mathbf{v}$ and performing a volume integration.

7.02.2.6 Boundary and Interface Conditions

7.02.2.6.1 General method

A boundary condition is a special case of an interface condition when certain properties are taken as known on one side of the interface. Sometimes, the properties are explicitly known (e.g., the three velocity components are zero on a no-slip surface), but often, an interface condition simply expresses the continuity of a conserved quantity. To obtain the continuity conditions for a quantity A, the general method is to start from the conservation equation of A in its integral form (see eqn [5]). We choose a cylindrical volume Ω (a pillbox) of infinitely small radius R where the top and bottom surfaces are located at a distance $\pm\epsilon$ from a discontinuity surface (see Figure 2). We choose two Cartesian axes Ox and Oy; we call $\mathbf{n}$ the upward

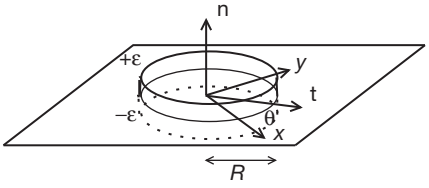


Figure 2 The pillbox volume used to derive the interface and boundary conditions.

unit vector normal to the interface and $\mathbf{t}$ a radial unit vector normal to the cylindrical side of the pillbox; and θ is the angle between $\mathbf{t}$ and Ox. If we now make the volume $\Omega(\epsilon)$ shrink to zero by decreasing ϵ at constant R, the volume integrals of the time-dependent and the source terms will also go to zero (unless the source term contains explicit surface terms like in the case of surface tension, but this is irrelevant at mantle scales). We simply get

$$\lim_{\epsilon\to 0}\int_{\Sigma(\epsilon)} \mathbf{J}_A\cdot d\mathbf{S} = 0 \qquad [62]$$

(the demonstration is here written for the vector flux $\mathbf{J}_A$ of the scalar A but is easily extended to tensor flux of a vectorial quantity $\mathbf{A}$). This integral can also be written as

$$\pi R^2[\mathbf{J}_A]\cdot\mathbf{n} + R\int_{-\epsilon}^{+\epsilon}\int_0^{2\pi}\mathbf{J}_A\cdot\mathbf{t}\,dz\,d\theta \qquad [63]$$

where $[X]$ is the jump of X across the interface, sometimes noted X^+–X^-. In most cases, the second term goes to zero with ϵ because the components of $\mathbf{J}_A$ are bounded or is exactly zero when the flux is not a function of θ (since the double integral becomes the product of an integral in z times $\int_0^{2\pi}\mathbf{t}\,d\theta = 0$). In these cases, the boundary condition for A becomes

$$[\mathbf{J}_A]\cdot\mathbf{n} = 0 \qquad [64]$$

At an interface, the normal flux of A must therefore be continuous. However, in some cases, for example, when $\mathbf{J}_A$ contains a z-derivative, the second term may not cancel and this happens in the case of boundaries associated with phase changes (see Section 7.02.2.6.3).

7.02.2.6.2 Interface conditions in the 1-D case and for bounded variables

Using the mass, momentum, energy, and entropy conservations in their conservative forms in eqns [9], [56], [57], and [40] and assuming for now that no variable becomes infinite at an interface, the interface conditions in the reference frame in which the interface is motionless are

$$\begin{aligned}
[\rho\mathbf{v}]\cdot\mathbf{n} &= 0\\
[\underline{\boldsymbol{\tau}}]\cdot\mathbf{n} - [P]\mathbf{n} &= 0\\
[\rho\mathbf{v}\mathcal{U} + \mathbf{q} - \underline{\boldsymbol{\tau}}\cdot\mathbf{v} + P\mathbf{v}]\cdot\mathbf{n} &= 0\\
\left[\rho\mathbf{v}\mathcal{S} + \frac{\mathbf{q}}{T}\right]\cdot\mathbf{n} &= 0
\end{aligned} \qquad [65]$$

(the gravitational force per unit mass and its potential are continuous). In these equations, we neglected the inertia and the kinetic energy terms in the second and third equations of eqn [65] as appropriate for the mantle. When these terms are accounted for (adding $[-\rho\mathbf{v}\otimes\mathbf{v}]\cdot\mathbf{n}$ to the second equation and $[-\rho\mathbf{v}\upsilon^2/2]\cdot\mathbf{n}$ to the third), these equations are known as the Hugoniot–Rankine conditions (Hugoniot, 1851–87; Rankine, 1820–72).

At the surface of a fluid, and on any impermeable interfaces where $\mathbf{v}\cdot\mathbf{n} = 0$ and $[\mathbf{v}] = 0$, the general jump conditions [65] without inertia imply that the heat flux $[\mathrm{q}]\cdot\mathbf{n}$, the entropy flux $[\mathbf{q}/\mathrm{T}]\cdot\mathbf{n}$ (and therefore the temperature T), and the stress components $[\underline{\boldsymbol{\tau}}]\cdot\mathbf{n} - [P]\mathbf{n}$ are continuous. In 3-D, four boundary conditions are necessary on a surface to solve for the three components of velocity and for the temperature. The temperature (or the heat flux) can be imposed, and, for the velocity and stress, either free-slip boundary conditions $\mathbf{v}\cdot\mathbf{n} = 0$, which is the first condition of eqn [65] and $(\underline{\boldsymbol{\tau}}\cdot\mathbf{n})\times\mathbf{n} = 0$, or no-slip boundary conditions (the horizontal surface velocity $\mathbf{v}$ is imposed) are generally used.

Inside the mantle, minerals undergo several phase transitions at depth and at least two of them, the olivine $\rightleftharpoons$ wadsleyite and the ringwoodite $\rightleftharpoons$ perovskite + ferropericlase transitions around 410 and 660 km depth, respectively, are sharp enough to be modeled by discontinuities. Conditions [65] have been used in several convection models (assuming therefore that all variables remain bounded).

When the kinetic energy is neglected, and the viscous stresses are much smaller than the pressure term, which are two approximations valid for the mantle, the last two boundary conditions are, assuming continuity of temperature,

$$\begin{aligned}
\left[\rho\mathbf{v}\left(\mathcal{U} + \frac{P}{\rho}\right)\right]\cdot\mathbf{n} + [\mathbf{q}]\cdot\mathbf{n} &= 0\\
[\rho\mathbf{v}\mathcal{S}]\cdot\mathbf{n} + \frac{1}{T}[\mathbf{q}]\cdot\mathbf{n} &= 0
\end{aligned} \qquad [66]$$

The diffusive flux $\mathbf{q}$ can be eliminated from these two equations. Since $p\mathbf{v}$ is continuous, and remembering that $\mathcal{U} + P/p$ is the enthalpy $\mathcal{H}$, we simply recognize the Clapeyron condition, which is latent heat release,

$$\Delta\mathcal{H} = T\Delta\mathcal{S} \qquad [67]$$

where the enthalpy and entropy jumps, $[\mathcal{H}]$ and $[\mathcal{S}]$, were replaced by their more traditional notations, $\Delta\mathcal{H}$ and $\Delta\mathcal{S}$. The heat flux is discontinuous across an interface,

$$\Delta\mathcal{H}\rho\mathbf{v}\cdot\mathbf{n} + [\mathbf{q}]\cdot\mathbf{n} = 0 \qquad [68]$$

and the discontinuity amounts to the enthalpy released by the mass flux that has undergone a chemical reaction or a phase change.

7.02.2.6.3 Phase change interfaces and stress continuity?

A mathematical dilemma arises in using the conditions [65] across a phase change interface that does not seem to have been considered except by Corrieu et al. (1995): a phase change induces a stress discontinuity at an interface.

We illustrate the mathematical difficulty by discussing the stress continuity in the second condition of eqn [65]. The problem arises from the term in $\partial\upsilon_z/\partial_z$ present in the rheological law [45] that becomes infinite when the material is forced to change its density discontinuously. To enforce the change in shape that occurs locally, the normal horizontal stresses have to become infinite, and therefore, their contributions to the force equilibrium of a pillbox do not vanish when the pillbox height is decreased.

To derive the appropriate interface condition, we have to consider again eqn [63] where $\mathbf{J}_A$ is substituted by $\underline{\sigma}$. The interface is considered as laterally infinite and perfectly flat (without curvature) and the viscosity is uniform. The only terms that may be unbounded on the interface are σ_{xx}, σ_{yy}, and σ_{zz}. Omitting the other stress components, which would make no contribution to the interface condition when ϵ goes to zero, the stress continuity becomes

$$\pi R^2 \left[\underline{\boldsymbol{\sigma}}\right]\cdot\mathbf{n} + R\mathbf{e}_x \int_{-\epsilon}^{+\epsilon}\int_0^{2\pi} \sigma_{xx}\cos\theta\,\mathrm{d}z\,\mathrm{d}\theta + R\mathbf{e}_y \int_{-\epsilon}^{+\epsilon}\int_0^{2\pi} \sigma_{yy}\sin\theta\,\mathrm{d}z\,\mathrm{d}\theta = 0 \quad [69]$$

Since R is small, we can replace the stresses on the cylindrical side of the pillbox by their first-order expansions, for example, $\sigma_{xx} = \sigma_{xx}(0) + (\partial\sigma_{xx}/\partial x)R\cos\theta + (\partial\sigma_{xx}/\partial y)R\sin\theta$. After integration in θ and division by πR^2, one gets

$$\left[\underline{\boldsymbol{\sigma}}\right]\cdot\mathbf{n} + \mathbf{e}_x\frac{\partial}{\partial x}\int_{-\epsilon}^{+\epsilon}\sigma_{xx}\,\mathrm{d}z + \mathbf{e}_y\frac{\partial}{\partial y}\int_{-\epsilon}^{+\epsilon}\sigma_{yy}\,\mathrm{d}z = 0 \quad [70]$$

This expression already demonstrates the continuity of σ_{zz}. Using $\sigma_{xx} = \sigma_{zz} + 2\eta\partial v_x/\partial x - 2\eta\partial v_z/\partial z$, and assuming that the viscosity remains uniform, we see that

$$\lim_{\epsilon\to 0}\int_{-\epsilon}^{\epsilon}\sigma_{xx}\,\mathrm{d}z = -2\eta\lim_{\epsilon\to 0}\int_{-\epsilon}^{\epsilon}\frac{\partial v_z}{\partial z}\mathrm{d}z = -2\eta[v_z] \quad [71]$$

The same result holds for the σ_{yy} term. Since v_z is discontinuous (because of mass-flux continuity and the density jump), a sudden change in volume implies a discontinuity of the tangential stresses. The stress boundary conditions are thus

$$[\tau_{xz}] - 2\eta\frac{\partial}{\partial x}[v_z] = [\tau_{yz}] - 2\eta\frac{\partial}{\partial y}[v_z] = [\tau_{zz} - P] = 0 \quad [72]$$

in disagreement with eqn [65]. Infinite horizontal stresses and discontinuous shear stresses occur in the presence of an interface with surface tension (Joseph and Renardy, 1993). In the case of a permeable and infinitely thin phase transition, there should also be a surface tension with a dynamic origin, simply related to $[v_z]$ when the viscosity is uniform. A rigorous mathematical demonstration that proves the existence of a normal stress discontinuity in the case of a curved interface in perfect analogy with usual surface tension and that is extended for a variable viscosity is given in Chambat et al. (2013). Notice also that the pressure and the viscous dissipation should also be unbounded. Until a complete clarification of these discontinuity problems is achieved, it seems that the usual continuity conditions across phase changes may not be mathematically consistent.

7.02.2.6.4 Weakly deformable surface of a convective cell

When a no-slip condition is imposed at the surface, both normal and shear stresses are present at the boundary. These stresses, according to the second interface condition [65], must balance the force $-\underline{\boldsymbol{\tau}}\cdot\mathbf{n} + P\mathbf{n}$ exerted by the fluid. This is reasonable for a laboratory experiment with a fluid totally enclosed in a tank whose walls are rigid enough to resist fluid traction. However, in the case of free-slip boundary conditions, it may seem strange that by imposing a zero vertical velocity, a finite normal stress results at the free surface. It is therefore worth discussing this point in more detail.

The natural boundary conditions should be that both the normal and the tangential stresses applied on the free deformable surface, $z = h(x, y, t)$, of a convective fluid are zero:

$$\left(\underline{\boldsymbol{\tau}}\cdot\mathbf{n} - P\mathbf{n}\right)_{\text{on } z=h} = 0 \quad [73]$$

(neglecting atmospheric pressure). In this expression, the topography h is unknown and the normal $\mathbf{n}$, computed at the surface of the planet, is $\mathbf{n} = (\mathbf{e}_z - \boldsymbol{\nabla}_H h)/\sqrt{1 + |\boldsymbol{\nabla}_H h|^2}$ where $\mathbf{e}_z$ is the unit vector along z, opposite to gravity, and $\boldsymbol{\nabla}_H$ the horizontal component of the gradient operator.

The variation of topography is related to the convective flow and satisfies

$$\frac{\partial h}{\partial t} + \mathbf{v}_H^0\cdot\boldsymbol{\nabla}_H h - v_z^0 = 0 \quad [74]$$

This equation expresses the fact that a material particle on the surface remains always on it. In this expression, $\mathbf{v}_H^0$ and v_z^0 are the horizontal and vertical velocity components at the surface of the planet. We will see in Section 7.02.4 that lateral pressure and stress variations are always very small compared to the average pressure (this is because in most fluids, and in the mantle, the lateral density variations remain negligible compared to the average density). This implies that the surface topography is not much affected by the internal dynamics and remains close to horizontal at long wavelengths, $|\boldsymbol{\nabla}_H h| \ll 1$. Boundary condition [73] and topography advection [74] can therefore be expanded to first order to give

$$\left(\underline{\boldsymbol{\tau}}\cdot\mathbf{e}_z - P\mathbf{e}_z\right)_{\text{on } z=0} \approx -\rho_0 g_0 h\mathbf{e}_z, \quad \frac{\partial h}{\partial t} = v_z^0 \quad [75]$$

where we again make use of the fact that the total stress remains close to hydrostatic, that is, $\mathbf{n}\cdot\underline{\boldsymbol{\tau}}\cdot\mathbf{n} \ll P$ (p_0 and g_0 are the surface values of density and gravity). To first order, the stress boundary condition on a weakly deformable top surface is therefore zero shear stress but with a time-dependent normal stress related to the surface topography and vertical velocity.

The convection equations with these boundary conditions could be solved, but this is not always useful. Since the boundary conditions involve both displacement h and velocity v_z^0, the solution is akin to an eigenvalue problem. It can be shown that for an internal density structure of wavelength, λ, v_z^0 goes to zero in a time of order $\overline{\eta}/\rho_0 g_0\lambda$ where $\overline{\eta}$ is the typical viscosity of the underlying liquid over the depth λ (Richards and Hager, 1984). For the Earth, this time is the characteristic time of postglacial rebound and is typically a few thousand years for wavelengths of a few thousand kilometers (e.g., Spada et al., 1992b).

For convection, where the characteristic times are much longer, it is thus appropriate to assume that the induced topography is in mechanical equilibrium with the internal density structure. A zero normal velocity can therefore be imposed and the resulting normal stress can be used to estimate the topography generated by the convective flow. Internal compositional interfaces can be treated in a similar manner if they are only weakly deformable (i.e., when their intrinsic density jumps are much larger than the thermal density variations). This is the case for the core–mantle boundary (CMB).

For rapid events (e.g., for a localized thermal anomaly impinging the Earth's surface), the time for topographic equilibration becomes comparable to the timescale of internal convective processes. In this case, the precise computation of a history-dependent topography is necessary and the finite elasticity of the lithosphere, the coldest part of the mantle, plays an important role (Zhong et al., 1996). Various studies have also shown that at short wavelength, the condition of free surface differs significantly from the more usual free-slip condition (Crameri et al., 2012; Kaus et al., 2010; Schmeling et al., 2008). In particular, one-sided subductions, as observed in nature and in analogical experiments, are reproduced much

more accurately by numerical codes that include a true deformable topography according to eqn [73].

7.02.3 Thermodynamic and Rheological Properties

Section 7.02.2 on conservation equations is valid for all fluids (although the interface conditions are mostly discussed when inertia and kinetic energy are negligible). The differences between mantle convection and core, oceanic, or atmospheric convection come from the thermodynamic and transport properties of solids that are very different from those of usual fluids. We review some basic general properties of solids in Section 7.02.3 and will be more specific in Section 7.02.6.

7.02.3.1 Equation of State and Solid Properties

The equation of state of any material (EoS) relates its pressure, density, and temperature. The equation of state of a perfect gas, $PV/T=$constant, is well known, but irrelevant for solids. Unfortunately, there is no equation for solids based on a simple and efficient theoretical model. In the mineralogical community, the third-order finite strain Birch–Murnaghan EoS seems highly favored (Birch, 1952). This equation is cumbersome and is essentially empirical. More physical approaches have been used in Vinet et al. (1987), Poirier and Tarantola (1998), and Stacey and Davis (2004), but it seems that for each solid, the EoS has to be obtained experimentally (for more details, see Stixrude, vol. 2).

In the simplest cases, the density varies around ρ_0 measured at temperature T_0 and pressure P_0 as

$$\rho=\rho_0\left(1-\alpha(T-T_0)+\frac{P-P_0}{K_T}\right) \qquad [76]$$

where the thermal expansivity α and incompressibility K_T have been defined in eqn [34]. This expression is a first-order expansion of any EoS. Equation [76] can however be misleading if one forgets that the parameters α and K_T cannot be constant but must be related through Maxwell relations (e.g., their definitions [34] imply that $\partial(\alpha\rho)/\partial P=-\partial(\rho/K_T)/\partial T$).

Equation [76] can be used for a very simple numerical estimate that illustrates an important characteristic of solid Earth geophysics. Typically for silicates $\alpha\approx10^{-5}\mathrm{K}^{-1}$, $K_T\approx10^{11}$ Pa, while temperature variations in the mantle, ΔT, are of a few 1000 K with a pressure increase between the surface and the core, ΔP, of order of 10^{11} Pa. This indicates that the overall density variations due to temperature differences are negligible compared to those due to pressure differences ($\alpha\Delta T\ll1$ but $\Delta P/K_T\approx1$). In planets, to first order, the radial density is only a function of pressure, not of temperature. This is opposite to most liquid or solid laboratory experiments, where the properties are usually controlled by temperature.

A very important quantity in the thermodynamics of solids is the Grüneisen parameter (Grüneisen, 1877–1949)

$$\Gamma=\frac{\sigma K_T}{\rho C_v}=\frac{1}{\rho C_v}\left(\frac{\partial P}{\partial T}\right)_V \qquad [77]$$

The Grüneisen parameter is dimensionless, does not vary much through the mantle (around 1), and can reasonably be considered as independent of the temperature. An empirical law (Anderson, 1979) relates Γ with the density,

$$\Gamma=\Gamma_0\left(\frac{\rho_0}{\rho}\right)^q \qquad [78]$$

where q is around 1. The Grüneisen parameter can also be related to the microscopic vibrational properties of crystals (Stacey, 1977). At high temperature, above the Debye temperature (Debye, 1884–1966), all solids have more or less the same heat capacity at constant volume. This is called the Dulong and Petit rule (Dulong, 1785–1838; Petit, 1791–1820). At high T, each atom vibrates and the thermal vibrational energy is equipartitioned between the three dimensions (degrees of freedom), which leads to $C_{Vm}=3R$ per mole of atoms, independent of the nature of the solid (R is the gas constant). Assuming that the mantle is made of pure forsterite Mg_2SiO_4 that contains seven atoms for a molar mass of 140 g, its heat capacity at constant volume is therefore close to $C_{Vm}=21R=174.56\ \mathrm{J\,K^{-1}\,mol^{-1}}$ or $C_V=1247\ \mathrm{J\,K^{-1}\,kg^{-1}}$. The approximate constancy of C_V and the fact that Γ is only a function of ρ allow us to integrate [77]

$$P=F(\rho)+\alpha_0K_T^0(T-T_0)\left(\frac{\rho}{\rho_0}\right)^{1-q} \qquad [79]$$

where α_0 and K_T^0 are the thermal expansivity and incompressibility at standard conditions and where $F(\rho)$ is a density-dependent integration constant. A rather simple but acceptable choice for the function $F(\rho)$, at least for mantle dynamicists, is the Murnaghan EoS (Murnaghan, 1951) at constant T that allows us to write an EoS for solids of the form

$$P=\frac{K_T^0}{n}\left[\left(\frac{\rho}{\rho_0}\right)^n-1\right]+\alpha_0K_T^0(T-T_0)\left(\frac{\rho}{\rho_0}\right)^{1-q} \qquad [80]$$

This equation could easily be used to derive any thermodynamic property like $\alpha(P,T)$ or $K_T(P,T)$. The exponent n expresses the variations of K_T with pressure at reference temperature T_0 as $K_T(P,T_0)=K_T^0+nP$. The value of n (also sometimes called K_0') is of order of 4 for most mantle materials (Stixrude and Lithgow-Bertelloni, 2005). This equation has been used implicitly in various models of mantle convection (e.g., Bercovici et al., 1989a, 1992; Glatzmaier, 1988). An important consequence of this EoS assuming $q\approx1$ is that αK_T is more or less constant and that

$$K_T\approx K_T^0\left(\frac{\rho}{\rho_0}\right)^n,\quad \alpha\approx\alpha_0\left(\frac{\rho}{\rho_0}\right)^{-n} \qquad [81]$$

In the mantle, the incompressibility increases and the thermal expansion decreases significantly with depth, typically by a factor ≈8. The geophysical consequences are further discussed in Section 7.02.6.5.1.

Two other thermodynamic equalities can also be straightforwardly deduced by chain rules of derivatives and will be used in the following. A relation between the two heat capacities C_P and C_V of the energy equations [38] can be derived from the two expressions for heat increments, eqn [31], and the definition of h, eqn [33],

$$C_P-C_V=\frac{\alpha T}{\rho}\left(\frac{\partial P}{\partial T}\right)_V \qquad [82]$$

The same expressions for heat increments, eqn [31], and the l and h definitions, eqn [33], imply that for an adiabatic transformation (when δQ and $\mathrm{d}\mathcal{S}$ are zero),

$$\left(\frac{\partial P}{\partial T}\right)_{\mathcal{S}} = \frac{\rho C_P}{\alpha T} \quad \text{and} \quad \left(\frac{\partial V}{\partial T}\right)_{\mathcal{S}} = -\frac{C_V}{\alpha K_T T} \qquad [83]$$

Equations [82] and [83] take simpler forms when the Grüneisen parameter [77] and the adiabatic compressibility defined by

$$K_{\mathcal{S}} = \rho\left(\frac{\partial P}{\partial \rho}\right)_{\mathcal{S}} = -\frac{1}{\rho}\left(\frac{\partial P}{\partial T}\right)_{\mathcal{S}} \Big/ \left(\frac{\partial V}{\partial T}\right)_{\mathcal{S}} \qquad [84]$$

are used; they are

$$\frac{C_P}{C_V} = \frac{K_{\mathcal{S}}}{K_T} = 1 + \Gamma\alpha T \qquad [85]$$

Since $\Gamma \approx 1$ and since $\alpha T \ll 1$, the two heat capacities are basically equal. The incompressibility K_S is defined similarly to K_T but at constant entropy. The theory of elastic waves introduces this parameter that can be obtained from seismic observations

$$K_{\mathcal{S}} = \rho\left(\upsilon_p^2 - \frac{4}{3}\upsilon_s^2\right) \qquad [86]$$

where υ_p and υ_s are the P and S wave velocities. We will see later (**Section 7.02.4.2.2**) that a vigorously convective fluid should be close to adiabatic. As this is the case for the mantle, K_S should be the incompressibility measured along the mantle density profile

$$K_{\mathcal{S}} = \rho\left(\frac{\partial P}{\partial \rho}\right)_{\mathcal{S}} = \rho\frac{\|\boldsymbol{\nabla} P\|}{\|\boldsymbol{\nabla}\rho\|} = \frac{\rho^2\|\mathbf{g}\|}{\|\boldsymbol{\nabla}\rho\|} \qquad [87]$$

The relations [86] and [87] connect a seismological observation $K_S(r)/\rho(r)$ to the density gradient of the real Earth $\rho(r)$ $g(r)/\|\boldsymbol{\nabla}\rho(r)\|$. This is the important Bullen hypothesis (Bullen, 1940) used to build the reference density of the Earth (e.g., Dziewonski and Anderson, 1981).

7.02.3.2 Rheology

In **Section 7.02.2.3** on momentum conservation, no assumption is made about the rheology of the fluid, that is, on the relation between the stress tensor and the flow itself. In contrast, the discussion of energy conservation (**Section 7.02.2.4**) relies on the assumption that the pressure-related work is entirely recoverable, eqn [32]; as a consequence, the work of the deviatoric stresses ends up entirely as a dissipative term, hence a source of entropy. In a real fluid, this may be wrong for two reasons: part of the deviatoric stresses may be recoverable and part of the isotropic work may not be recoverable. In the first case, elasticity may be present, and in the second case, bulk viscosity.

7.02.3.2.1 Elasticity

On a very short timescale, the mantle is an elastic solid in which compressional and shear waves propagate (e.g., Kennett, 2001). In an elastic solid, the linear strain tensor,

$$\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}} = \frac{1}{2}\left(\boldsymbol{\nabla}\mathbf{u} + [\boldsymbol{\nabla}\mathbf{u}]^t\right) \qquad [88]$$

in which $\mathbf{u}$ is the displacement vector (this is valid for small deformations, see e.g., Malvern, 1969; Landau and Lifchitz, 2000, for the large deformation case), is linearly related to the stress tensor,

$$\sigma_{ij}^{\mathrm{e}} = \sum_{k,l} A_{ijkl}^{\mathrm{e}} \epsilon_{kl}^{\mathrm{e}} \qquad [89]$$

where A^{e} is the fourth-rank stiffness tensor. Since both the stress and the strain tensors are symmetrical and because of the Maxwell thermodynamic relations for internal energy (including the elastic energy), $\partial^2\mathcal{U}/\partial\epsilon_{ij}\partial\epsilon_{kl} = \partial^2\mathcal{U}/\partial\epsilon_{kl}\partial\epsilon_{ij}$, the elastic tensor is invariant to permutations of i and j, k and l, and ij and kl. This leaves in the most general case of anisotropy 21 independent stiffness coefficients (Malvern, 1969). In crystals, this number decreases with the number of symmetries of the unit cell. For isotropic elastic solids, only two parameters are needed, the incompressibility K and the rigidity μ_{R}, and the elastic behavior satisfies

$$\underline{\boldsymbol{\sigma}}^{\mathrm{e}} = K\mathrm{tr}\left(\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}}\right)\underline{\mathbf{I}} + 2\mu_{\mathrm{R}}\left(\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}} - \frac{1}{3}\mathrm{tr}\left(\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}}\right)\underline{\mathbf{I}}\right) \qquad [90]$$

where $\mathrm{tr}\left(\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}}\right) = \boldsymbol{\nabla}\cdot\mathbf{u}$.

Two remarks can be made on this rapid presentation of elasticity that are more deeply developed in textbooks of mechanics (e.g., Landau and Lifchitz, 2000; Malvern, 1969) or of seismology (e.g., Dahlen and Tromp, 1998). First, the expression [90] assumes that the displacement vector is computed from an initial situation where the solid is perfectly stress-free, that is, $\underline{\boldsymbol{\sigma}}^{\mathrm{e}} = 0$ when $\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}} = 0$. In practical problems, only incremental displacements with respect to an initial prestressed state are known and $\underline{\boldsymbol{\sigma}}^{\mathrm{e}}$ has to be understood as a variation of the stress tensor. Second, temperature variations are associated with changes in elastic stresses and the definition of incompressibility K should take these variations into account. The incompressibility should be $K = K_S$ for rapid adiabatic seismic waves and $K = K_T$ for isothermal variations. The other elastic parameters that are often introduced, Poisson's ratio, Young's modulus (Young, 1773–1829), and Lamé's parameters (Lamé, 1795–1870), are simple functions of incompressibility and rigidity. Since the term proportional to μ_{R} is traceless, equation [90] leads to $\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\mathrm{e}}\right) = 3K\mathrm{tr}\left(\underline{\dot{\boldsymbol{\varepsilon}}}^{\mathrm{e}}\right)$, and the rheology law can also be written in terms of compliance (i.e., getting $\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$ as a function of $\underline{\boldsymbol{\sigma}}^{\mathrm{e}}$),

$$\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}} = \frac{1}{9K}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\mathrm{e}}\right)\underline{\mathbf{I}} + \frac{1}{2\mu_{\mathrm{R}}}\left(\underline{\boldsymbol{\sigma}}^{\mathrm{e}} - \frac{1}{3}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\mathrm{e}}\right)\underline{\mathbf{I}}\right) \qquad [91]$$

In these equations, the trace of the stress tensor can also be replaced by the pressure definition

$$\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\mathrm{e}}\right) = -3P \qquad [92]$$

The momentum equation [15] remains valid in a purely elastic solid (except that the advective transport is generally neglected, $\mathrm{D}/\mathrm{D}t \approx \partial/\partial t$), but the discussion of energy conservation and thermodynamics is different for elastic and viscous bodies. The work of the elastic stress is entirely recoverable: a deformed elastic body returns to its undeformed shape when the external forces are released. The internal energy change due to the storage of elastic stress is $\delta W = V\underline{\boldsymbol{\sigma}}_{\mathrm{e}} : \mathrm{d}\underline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$ instead of

$\delta W = -P\mathrm{d}V$, and this is provided by the deformation work term $\underline{\boldsymbol{\tau}} : \nabla \mathrm{v}$, which is therefore nondissipative. Thus, for an elastic body, the temperature equations [38] and the entropy equation [39] hold, but with the $\underline{\boldsymbol{\tau}} : \nabla \mathrm{v}$, source term removed.

7.02.3.2.2 Viscous Newtonian rheology

On a very long timescale, it is reasonable to assume that the internal deviatoric stresses become eventually relaxed and dissipated as heat. This is the assumption that we have implicitly made and that is usual in fluid mechanics. Since the dissipative term is $\underline{\boldsymbol{\tau}} : \nabla \mathrm{v} = \underline{\boldsymbol{\tau}} : \underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon}$ and must be positive according to the second law, this suggests a relationship between velocity-related stresses and velocity derivatives such that the total stress tensor has the form

$$\sigma_{ij}^{\upsilon} = -P\delta_{ij} + A_{ijkl}^{\upsilon}\dot{\varepsilon}_{kl}^{\upsilon} \qquad [93]$$

δ_{ij} being the Kronecker symbol (Kronecker, 1823–91). Except for the time derivative, the only formal difference between this expression and eqn [89] is that pressure exists in a motionless fluid but is always associated with deformation in an elastic solid.

Using the same arguments as for the elastic case, the viscous rheology in the isotropic case can therefore be written in terms of stiffness

$$\underline{\boldsymbol{\sigma}}^{\upsilon} = \left(-P + \zeta \mathrm{tr}\left(\underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon}\right)\right)\underline{\mathbf{I}} + 2\eta\left(\underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon} - \frac{1}{3}\mathrm{tr}\left(\underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon}\right)\underline{\mathbf{I}}\right) \qquad [94]$$

where $\mathrm{tr}\left(\underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon}\right) = \nabla\cdot\mathbf{v}$. Using $\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\upsilon}\right) = 3\left(-P + \zeta \mathrm{tr}\left(\underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon}\right)\right)$, the rheology can also be expressed in terms of compliance

$$\underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon} = \frac{1}{9\zeta}\left(3P + \mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\upsilon}\right)\right)\underline{\mathbf{I}} + \frac{1}{2\eta}\left(\underline{\boldsymbol{\sigma}}^{\upsilon} - \frac{1}{3}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\upsilon^2}\right)\underline{\mathbf{I}}\right) \qquad [95]$$

The two parameters η and ζ are positive according to the second law and are called the shear and bulk viscosities. When they are intrinsic material properties (i.e., independent of the flow itself), the fluid is called linear or Newtonian. The hypothesis of isotropy of the rheology is probably wrong for a mantle composed of highly anisotropic materials (see Karato, 1998), but only a few papers have tried to tackle the problem of anisotropic viscosity (Christensen, 1997a; Muhlhaus et al., 2004; Pouilloux et al., 2007).

Since $\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}^{\upsilon}\right)/3 = -P + \zeta\,\nabla\cdot\mathbf{v}$, the isotropic average of the total stress is not the pressure term, unless $\zeta\nabla\cdot\mathbf{v} = 0$. Therefore, part of the stress work, $\underline{\boldsymbol{\tau}} : \underline{\dot{\boldsymbol{\varepsilon}}}^{\upsilon}$, during isotropic compaction could be dissipated in the form of the heat source $\zeta(\nabla\cdot\mathbf{v})^2$. A density-independent bulk viscosity allows an infinite compression under a finite isotropic stress. The bulk viscosity parameter ζ is generally only introduced to be immediately omitted and we will do the same. However, using eqn [94] with $\zeta = 0$ but keeping $\nabla\cdot\mathbf{v} \neq 0$ is not valid since it would remove all resistance to isotropic compression. We will see that the correct physical behavior is obtained by setting $\zeta = +\infty$: elastic stresses are present and provide the resistance to isotropic compression. A finite bulk viscosity, or some equivalent concept, is however necessary to handle two-phase compaction problems (Bercovici et al., 2001a; McKenzie, 1984) (see **Section 7.02.5.2**).

7.02.3.2.3 Maxwellian viscoelasticity

To account for the fact that the Earth behaves elastically on short time constants and viscously at long times, it is often assumed that under the same stress, the deformation has both elastic and viscous components. By summing the viscous compliance equation [95] with the time derivative of the elastic compliance equation, [91] and in the case of an infinite bulk viscosity ζ, we get

$$\underline{\dot{\boldsymbol{\varepsilon}}} = \frac{1}{9K}\mathrm{tr}\left(\underline{\dot{\boldsymbol{\sigma}}}\right)\underline{\mathbf{I}} + \frac{1}{2\eta}\left(\underline{\boldsymbol{\sigma}} - \frac{1}{3}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}\right)\underline{\mathbf{I}}\right) + \frac{1}{2\mu_{\mathrm{R}}}\left(\underline{\dot{\boldsymbol{\sigma}}} - \frac{1}{3}\mathrm{tr}\left(\underline{\dot{\boldsymbol{\sigma}}}\right)\underline{\mathbf{I}}\right) \qquad [96]$$

where $\underline{\boldsymbol{\sigma}} = \underline{\boldsymbol{\sigma}}^{\upsilon} = \underline{\boldsymbol{\sigma}}^{\mathrm{e}}$ and $\underline{\boldsymbol{\varepsilon}} = \underline{\boldsymbol{\varepsilon}}^{\upsilon} + \underline{\boldsymbol{\varepsilon}}^{\mathrm{e}}$. This time-dependent rheological law is the constitutive law of a linear Maxwell solid.

A few simple illustrations of the behavior of a Maxwellian body will illustrate the physical meaning of equation [96] (see also Ribe, this volume). First, we can consider the case where stress and strain are simple time-dependent sinusoidal functions with frequency ω (i.e., $\underline{\boldsymbol{\sigma}} = \underline{\boldsymbol{\sigma}}_0\exp(\mathrm{i}\omega t)$ and $\underline{\boldsymbol{\varepsilon}} = \underline{\boldsymbol{\varepsilon}}_0\exp(i\omega t)$). The solution to this problem can then be used to solve other time-dependent problems by the Fourier or Laplace transforms. Equation [96] becomes

$$\underline{\boldsymbol{\varepsilon}}_0 = \frac{1}{9K}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}_0\right)\underline{\mathbf{I}} + \frac{1}{2\mu_{\mathrm{R}}}\left(1 - \frac{i}{\omega\tau}\right)\left(\underline{\boldsymbol{\sigma}}_0 - \frac{1}{3}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}_0\right)\underline{\mathbf{I}}\right) \qquad [97]$$

where $\tau = \eta/\mu_{\mathrm{R}}$ is the Maxwell time, eqn [1]. This equation can be compared to eqn [91] and shows that the solution of a viscoelastic problem is formally equivalent to that of an elastic problem with a complex elastic rigidity. This is called the correspondence principle.

We can also solve the problem of a purely 1-D Maxwellian body (only σ_{zz} and ϵ_{zz} are nonzero), submitted to a sudden load $\sigma_{zz} = \sigma_0 H(t)$ (where H is the Heaviside distribution (Heaviside, 1850–1925)) or to a sudden strain $\epsilon_{zz} = \epsilon_0 H(\mathrm{t})$. The solutions are, for $t \geq 0$,

$$\epsilon_0 = \frac{1}{3k\mu_{\mathrm{R}}}\sigma_0 + \frac{1}{3\eta}\sigma_0 t \qquad [98]$$

and

$$\sigma_0 = 3k\mu_{\mathrm{R}}\exp\left(-k\frac{t}{\tau}\right)\epsilon_0 \qquad [99]$$

respectively, with $k = 3K/(3K + \mu_{\mathrm{R}})$. In both cases, the instantaneous elastic response is followed by a viscous flow. In the first case, the finite elastic deformation is followed by a steady flow. In the second case, the initial elastic stresses are then dissipated by viscous relaxation over a time constant, τ/k. This time constant is different from the Maxwell time constant as both deviatoric and nondeviatoric stresses are present. For mantle material, the time τ/k would however be of the same order as the Maxwell time constant τ (in the midmantle, $K \approx 2\mu_{\mathrm{R}} \approx 200$ GPa).

From equation [96], we can now understand what rheology must be used for a compressible viscous mantle. For phenomena that occur on time constants much larger than the Maxwell time, the deviatoric stresses can only be supported by the viscosity. As a typical viscosity for the deep mantle is in the range 10^{19}–10^{22} Pa s (see **Sections 7.02.4** and **7.02.6**), the appropriate Maxwell times are in the range 30 year–30 kyr, much shorter than those of convection. By contrast, the isotropic stress remains only supported by elasticity in the approximation where the bulk viscosity ζ is infinitely large. The appropriate rheology for mantle convection is therefore given by

$$\underline{\dot{\boldsymbol{\varepsilon}}} = \frac{1}{9K}\mathrm{tr}\left(\underline{\dot{\boldsymbol{\sigma}}}\right)\underline{\mathbf{I}} + \frac{1}{2\eta}\left(\underline{\boldsymbol{\sigma}} - \frac{1}{3}\mathrm{tr}\left(\underline{\boldsymbol{\sigma}}\right)\underline{\mathbf{I}}\right) \qquad [100]$$

This equation is simultaneously a rheology equation for the deviatoric stress and an EoS for the isotropic stress. Using $P=-\mathrm{tr}(\boldsymbol{\sigma})/3$, the stress tensor verifies

$$\underline{\boldsymbol{\sigma}}=-P\underline{\mathbf{I}}+2\eta\left(\underline{\dot{\boldsymbol{\varepsilon}}}-\frac{\dot{P}}{3K}\underline{\mathbf{I}}\right) \quad [101]$$

This equation is intrinsically a viscoelastic equation that can be replaced by a purely viscous equation plus an EoS:

$$\underline{\boldsymbol{\sigma}}=-P\underline{\mathbf{I}}+2\eta\left(\underline{\dot{\boldsymbol{\varepsilon}}}-\frac{1}{3}\mathrm{tr}\left(\underline{\dot{\boldsymbol{\varepsilon}}}\right)\underline{\mathbf{I}}\right) \quad [102]$$

$$\mathrm{tr}\left(\underline{\boldsymbol{\varepsilon}}\right)=\frac{P}{K} \quad [103]$$

Equation [102] is therefore the appropriate limit of eqn [94] for slow deformation, when $\zeta=+\infty$ and when isotropic compaction is resisted by the elastic stresses.

The use of a Maxwellian viscoelastic body to represent the mantle rheology on short timescale remains however rather arbitrary. Instead of summing the elastic and viscous deformations for the same stress tensor, another linear viscoelastic body could be obtained by partitioning the total stress into elastic and viscous components for the same strain rate. Basically, instead of having the elasticity and the viscosity added like a spring and a dashpot in series (Maxwell rheology), this Kelvin–Voigt rheology would connect in parallel a viscous dashpot with an elastic spring (Kelvin, 1824–1907; Voigt, 1850–1919). Of course, further degrees of complexity could be reached by summing Maxwell and Voigt bodies, in series or in parallel. Such models have sometimes been used for the Earth, but the data that could support or dismiss them were scarce (Ross and Schubert, 1986; Yuen et al., 1986). However, the unprecedented precision of the observations of large-scale deformations that followed the magnitude 9 seisms of Indonesia (Aceh earthquake, 2004) and Japan (Tohoku earthquake, 2011) should soon allow to reevaluate and constrain the transient rheology of the mantle (e.g., Trubienko et al., 2013).

7.02.3.2.4 Nonlinear rheologies

Even without elasticity and bulk viscosity, the assumption of a linear Newtonian rheology for the mantle is problematic. The shear viscosity cannot be a direct function of velocity since this would contradict the necessary Galilean invariance of material properties. However, since the shear viscosity is a scalar, it could be any function of the scalar invariants of the stress tensor. There are three invariants of the stress tensor: its trace (related to the pressure), its determinant, and the second invariant $I_2=\left(\underline{\boldsymbol{\tau}}:\underline{\boldsymbol{\tau}}\right)/2$ (where as in eqn [30], the double dots denote tensor contraction). The viscosity could therefore be a function of $\det\left(\underline{\boldsymbol{\tau}}\right)$ and I_2 in addition to the thermodynamic variables P and T and the mineralogical variables, for example, composition, grain size, and fabrics. Because viscosity must be a positive quantity, its stress dependence is assumed to be related to I_2 (which is always positive) and not to $\det\left(\underline{\boldsymbol{\tau}}\right)$ (which can change sign).

The main mechanism of solid-state deformation pertinent to mantle conditions (excluding the brittle and plastic deformations) is either diffusion creep or dislocation creep (see Poirier, 1991). In the first case, finite deformation is obtained by summing the migrations of individual atoms exchanging their positions with crystalline lattice vacancies. In crystals, the average number of lattice vacancies C varies with pressure, P, and temperature, T, according to Boltzmann statistics (Boltzmann, 1844–1906),

$$C\propto\exp\left(-\frac{PV}{RT}\right) \quad [104]$$

(V is the atomic volume and R the gas constant). A mineral is composed of grains of size d with an average concentration of lattice vacancies C_0. Submitted to a deviatoric stress τ, a gradient of vacancies of order $|\nabla C|\propto(C_0/d)(\tau V/RT)$ appears due to the difference in stress regime between the faces in compression and the faces in extension ($\tau V\ll RT$) (see Poirier, 1991; Ranalli, 1995; Schubert et al., 2001; Turcotte and Schubert, 1982). This induces the flux of atoms (number of atoms per unit surface and unit time):

$$J\propto D\frac{C_0}{d}\frac{\tau}{RT} \quad [105]$$

where D is a diffusion coefficient. This flux of atoms goes from the grain faces in compression to the grain faces in extension. Along the direction of maximum compression, each crystal grain shortens by a quantity δd, which corresponds to a total transport of $d^2\delta d/V$ atoms. These atoms can be transported in a time δt by the flux J_V across the grain of section d^2 (with volume diffusion D_V). They can also be transported by grain boundary flux J_B (with grain boundary diffusion D_B) along the grain interfaces through a surface hd (h being the thickness of the grain boundary), according to

$$\frac{d^2\delta d}{V}\approx J_D d^2\delta t,\quad \text{or}\quad \frac{d^2\delta d}{V}\approx J_B dh\delta t \quad [106]$$

As $\dot{\boldsymbol{\varepsilon}}=(\delta d/\delta t)/d$, the previous equations lead to the stress–strain rate relationship:

$$\underline{\dot{\boldsymbol{\varepsilon}}}\propto\frac{C_0V}{d^2RT}\left(D_V+D_B\frac{h}{d}\right)\underline{\boldsymbol{\tau}} \quad [107]$$

These diffusion mechanisms lead to a Newtonian rheology but with a grain size dependence of the viscosity: $\eta\propto d^2$ for Nabarro–Herring creep with diffusion inside the grain (Nabarro, 1916–2006; Herring, 1914–2009) and $\eta\propto d^3$ for Coble grain boundary creep (Coble, 1928–92). The viscosity is also very strongly T-dependent not so much because of the explicit factor T in eqn [107], but because diffusion is a thermally activated process, $D\propto\exp(-E_{dif}/RT)$, where E_{dif} is an activation enthalpy of diffusion. As the diffusion increases with temperature, the viscosity in the diffusion creep regime decreases with temperature.

In the case of dislocation creep, lines or planar imperfections are present in the crystalline lattice, and macroscopic deformation occurs by collective slip motion along these imperfections, called dislocations. Instead of the grain size d for diffusion creep, the mean spacing d_d between dislocations provides the length scale. This distance is experimentally found to decrease with the intensity of the deviatoric stresses and to vary as $\approx 1/\sqrt{I_2}$. Therefore, instead of a diffusion creep with a viscosity in d^n, the resulting rheology is rather in $I_2^{-n/2}$ and is also thermally activated with an activation energy E_{dis}. Dislocation creep leads to a nonlinear regime where the equivalent

viscosity varies with the second invariant with a power $-n/2$, where n is typically of order 2,

$$\underline{\dot{\varepsilon}} \propto I_2^{n/2} \exp(-E_{\text{dis}}/RT)\underline{\tau} \qquad [108]$$

and to a viscosity also strongly decreasing with temperature (this relationship is often written, in short, $\underline{\dot{\varepsilon}} \propto \underline{\tau}^m$ with a stress exponent m of order 3 but $\underline{\tau}^m$ really means $I_2^{(m-1)/2}\underline{\tau}$).

In general, for a given stress and a given temperature, the mechanism with the smallest viscosity (largest strain rate) prevails. Whether linear (grain size-dependent) or nonlinear (stress-dependent), viscosities are also strongly dependent upon temperature, pressure, melt content, water content, mineralogical phase, and oxygen fugacity (e.g., Hirth and Kolhstedt, 1996). In **Section 7.02.6.3**, we will further discuss the rheological mechanisms appropriate for the Earth.

7.02.4 Physics of Convection

The complex and very general system of equations that we have discussed in **Sections 7.02.2** and **7.02.3** can be used to model an infinite number of mantle flow situations. Mantle flow can sometimes be simply modeled as driven by the motion of plates (some examples are discussed in Zhong et al., Parmentier, and King, this volume). It can also be induced by compositional density anomalies (some examples are discussed in Tackley, this volume, and Forte, vol. 3). However, a fundamental cause of motion is due to the interplay between density, pressure, and temperature, and this is called thermal convection.

The phenomenon of thermal convection is common to all fluids (gas, liquid, and creeping solids), and it can be illustrated by simple experiments (see **Chapter 7.03**). The simplest can be done using water and an experimental setup called the shadowgraph method. Parallel light enters a transparent fluid put in a glass tank and is deflected where there are refractive index gradients due to temperature variations in the fluid. A pattern of bright regions and dark shadows is formed on a screen put on the other side of the tank. From this shadowgraph, the structure of the temperature pattern can be qualitatively assessed (see examples of shadowgraphs in Tritton, 1988).

7.02.4.1 Basic Balance

From a simple thought experiment on thermal convection, we can derive the basic dynamic balance of convection. Let us consider a volume of fluid, Ω, of characteristic size a, in which there is a temperature excess ΔT with respect to the surrounding fluid. The fluid is subject to a gravity $\mathbf{g}$, and it has an average density ρ and thermal expansivity α. The volume Ω, because of its anomalous temperature, experiences an Archimedean force, or buoyancy (Archimedes c.287–212 BC), given by

$$\mathbf{f} = -c_1 a^3 \rho \alpha \Delta T \mathbf{g} \qquad [109]$$

(c_1 is a constant taking into account the shape of Ω, e.g., $c_1 = 4\pi/3$ for a sphere of radius a). If the volume Ω is in a fluid of viscosity η, it will sink or rise with a velocity given by the Stokes law (Stokes, 1819–1903):

$$\mathbf{v}_s = -c_1 c_2 \frac{a^2 \rho \alpha \Delta T \mathbf{g}}{\eta} \qquad [110]$$

(c_2 is a drag coefficient accounting for the shape of Ω, i.e., $c_2 = 1/(6\pi)$ for a sphere).

During its motion, the volume Ω exchanges heat by diffusion with the rest of the fluid and the diffusion equation [44] tells us that a time of order

$$t_d = c_3 \frac{\rho C_p a^2}{k} \qquad [111]$$

is needed before temperature equilibration with its surroundings. During this time, the fluid parcel travels the distance $l = v_s t_d$.

A natural indication of the possibility that the parcel of fluid moves can be obtained by comparing the distance l to the characteristic size a. When $l \gg a$, that is, when the fluid volume can be displaced by several times its size, motion will be possible. On the contrary when $l \ll a$, thermal equilibration will be so rapid that no motion will occur.

The condition $l \gg a$, when v_s and t_d are replaced by the previously mentioned expressions, depends on only one quantity, the Rayleigh number Ra,

$$Ra = \frac{\rho^2 \alpha \Delta T g a^3 C_P}{\eta k} = \frac{\alpha \Delta T g a^3}{\kappa \nu} \qquad [112]$$

in terms of which motion occurs when $Ra \gg 1$ (assuming that $c_1 c_2 c_3 \approx 1$). The Rayleigh number compares the driving mechanism (e.g., the Archimedean buoyancy) to the two resistive mechanisms, the diffusion of heat, represented by κ (see eqn [43]), and the diffusion of momentum, represented by ν (see eqn [50]).

This simple balance suggests that a large nondimensional number Ra favors fluid motion. How large Ra needs to be is a question that we cannot address at this moment, but it will be discussed in **Section 7.02.4.4**. Convection lifts hot fluid and causes cold fluid to sink (assuming $\alpha > 0$, which is true for most fluids and for the mantle). A convective system will rapidly reach an equilibrium where all thermal heterogeneities are swept up or down (if Ra is large) or thermally equilibrated (if Ra is small), unless a forcing mechanism continuously injects new cold parcels at the top and new hot parcels at the bottom. This can be done by cooling the top surface or heating the bottom one. When a fluid is heated from the side, a lateral temperature anomaly is constantly imposed and the liquid lateral thermal equilibration is prevented. The fluid remains in motion regardless of the amplitude of the imposed temperature anomaly.

7.02.4.2 Two Simple Solutions

7.02.4.2.1 The diffusive solution

Trying to directly and exactly solve the mass, momentum, energy, and Poisson's equations and accounting for a realistic EoS would certainly be a formidable task. This complex system of equations has however two rather obvious but opposite solutions.

A steady and motionless solution is indeed possible. The assumption $\partial/\partial t = 0$ and $\mathbf{v} = 0$ satisfies the mass equation [9], the momentum equation [16] when the pressure is hydrostatic,

$$0 = -\nabla P + \rho \mathbf{g} \qquad [113]$$

and the energy equation [38] when the temperature is diffusive (using the Fourier law [42]),

$$\nabla \cdot (k \nabla T) + \rho H = 0 \qquad [114]$$

Solving analytically for the hydrostatic pressure and the diffusive temperature is trivial when H, k, $\mathbf{g}$, and ρ are uniform. For example, choosing a depth z positive downward, we get

$$P = \rho g z, \quad T = T_0 + \Delta T \frac{z}{h} + \frac{1}{2} \rho H z (h - z) \qquad [115]$$

across a conductive solution with $T = T_0$, $P = 0$ at $z = 0$, and $T = T_0 + \Delta T$ at $z = h$. Computing analytically the conductive solution in the mantle remains feasible but could be quite cumbersome if one introduces a realistic EoS and computes gravity in agreement with the density distribution using Poisson's equation [51]. In Section 7.02.4.4, we will understand why the fluid does not necessarily choose the diffusive solution.

7.02.4.2.2 The adiabatic solution

The previous diffusive solution was obtained for a steady motionless situation. However, the opposite situation where the velocities are very large also corresponds to a rather simple situation. The energy equation [38] can also be written as

$$\rho C_V T \left(\frac{\mathrm{D} \ln T}{\mathrm{D}t} - \Gamma \frac{\mathrm{D} \ln \rho}{\mathrm{D}t} \right) = -\nabla \cdot \mathbf{q} + \underline{\boldsymbol{\tau}} : \nabla \mathbf{v} + \rho H \qquad [116]$$

or

$$\rho C_P T \left(\frac{\mathrm{D} \ln T}{\mathrm{D}t} - \frac{\alpha}{\rho C_P} \frac{\mathrm{D}P}{\mathrm{D}t} \right) = -\nabla \cdot \mathbf{q} + \underline{\boldsymbol{\tau}} : \nabla \mathbf{v} + \rho H \qquad [117]$$

The right sides of these equations were previously shown to be equal to $\rho T \mathrm{D}S/\mathrm{D}t$ in eqn [39]. If we decrease the viscosity in a fluid, the convective velocity increases. The advection terms $\mathbf{v} \cdot \nabla T$, $\mathbf{v} \cdot \nabla \ln \rho$ and $\mathbf{v} \cdot \nabla P$ become then much larger than the time-dependent, diffusion, and radioactive production terms. With a low viscosity, the fluid becomes also unable to sustain large stresses. As a consequence, when convection is vigorous enough, the fluid should evolve toward a situation where

$$\begin{aligned} (\nabla \ln T)_S - \Gamma (\nabla \ln \rho)_S &= 0 \\ (\nabla \ln T)_S - \frac{\alpha}{\rho C_P} (\nabla P)_S &= 0 \end{aligned} \qquad [118]$$

Since such equations imply that the entropy is exactly conserved, $\mathrm{D}S/\mathrm{D}t = 0$, this equilibrium is called the adiabatic equilibrium. We added a subscript $[\ ]_S$ to denote the isentropic state.

Notice that, since the Grüneisen parameter is only density-dependent (see eqn [78]), density and temperature are simply related along the adiabat. For example, if the Grüneisen parameter is a constant, Γ_0 (using $q = 0$ in eqn [78]), the first of the equations [118] implies

$$T = T_0 \left(\frac{\rho}{\rho_0} \right)^{\Gamma 0} \qquad [119]$$

where T_0 and ρ_0 are two reference values. This equation implies that the adiabatic temperature increases by a factor 1.72 (e.g., from 1300 to 2230 K) from the asthenosphere ($\rho_0 \approx 3200\ \mathrm{kg\ m^{-3}}$) to the CMB ($\rho \approx 5500\ \mathrm{kg\ m^{-3}}$), if we assume $\Gamma_0 = 1$.

7.02.4.2.3 Stability of the adiabatic gradient

When a fluid is compressed, it heats up and it cools down when decompressed. This is the same physics that explains why atmospheric temperature decreases with altitude. Of course, this adiabatic effect is vanishingly small in laboratory experiments, but not always in nature.

If a parcel of fluid is rapidly moved up or down along z by a distance a, it changes its temperature adiabatically, by the quantity $a(\mathrm{d}T/\mathrm{d}z)_S$. However, the surrounding fluid will be at the temperature $a(\mathrm{d}T/\mathrm{d}z)$ where $\mathrm{d}T/\mathrm{d}z$ is just the temperature gradient, not necessarily adiabatic, of the fluid at rest. We can define ΔT_{na} as the nonadiabatic temperature $\Delta T_{\mathrm{na}} = a(\mathrm{d}T/\mathrm{d}z - (\mathrm{d}T/\mathrm{d}z)_S)$. The parcel being warmer or colder than the surroundings will rise or sink with a Stokes velocity that, rather than eqn [110], will be of order

$$\mathbf{v}_s = -c_1 c_2 \frac{a^2 \rho \alpha \Delta T_{\mathrm{na}} \mathbf{g}}{\eta} \qquad [120]$$

Since z is depth and $\mathrm{d}z$ is positive along $\mathbf{g}$, the adiabatic gradient is positive and the fluid parcel is locally unstable when the gradient in the surrounding fluid is larger (superadiabatic) than the adiabatic gradient. On the contrary, a subadiabatic gradient is stable with respect to convection. It is therefore not the total temperature difference between the top and the bottom of the fluid that drives motion, but only its nonadiabatic part.

To compare the Stokes velocity with the thermal equilibration time, we need to introduce a modified Rayleigh number

$$Ra = \frac{\alpha \Delta T_{\mathrm{na}} g_0 a^3}{\kappa \nu} \qquad [121]$$

This number is based on the nonadiabatic temperature difference in excess of the adiabatic variation imposed over the height a.

We have shown that inside a convective cell, the thermal gradient should be superadiabatic. Superadiabaticity is the source of convective instability, but vigorous convective stirring involves largely rapid adiabatic vertical motion; so much of convecting fluid is indeed adiabatic, while most of superadiabaticity is bound up in the thermal boundary layers where the vertical motion goes to zero. This mechanism suggests that an adiabatic reference background should not be such a bad assumption for the bulk of a convective fluid.

This adiabaticity hypothesis should, however, not be taken too literally (Jeanloz and Morris, 1987). In most numerical simulations, the resulting averaged geotherm can be far (a few hundred kelvins) from adiabatic (Bunge et al., 2001). First, radioactive heating, dissipation, and diffusion are never totally negligible; second, even if each fluid parcel follows its own adiabatic geotherm, the average geotherm may not correspond to any particular adiabat.

7.02.4.3 Approximate Equations

7.02.4.3.1 Depth-dependent reference profiles

We assume that the thermodynamic state is not far from a hydrostatic adiabat; thus, we choose this state as a reference.

We denote all the reference variables with an overbar, the hydrostatic pressure given by

$$\nabla\overline{P} = \overline{\rho}\,\mathbf{g} \qquad [122]$$

and adiabatic temperature and densities obeying eqn [118]:

$$\begin{aligned}\nabla\overline{T} &= \frac{\overline{\alpha}\,\overline{\mathbf{g}}}{\overline{C_p}}\overline{T}\\ \nabla\overline{\rho} &= \frac{\overline{\alpha}\,\overline{\mathbf{g}}}{\overline{C_p}\Gamma}\overline{\rho}\end{aligned} \qquad [123]$$

where all the parameters are computed along the reference geotherm and where $\overline{\mathbf{g}}$ has been solved using Poisson's equation [51] with the reference density $\overline{\rho}$.

The reference parameters are depth-dependent and usually, even for a simple EoS, cannot be analytically obtained. They can however be computed numerically from eqn [118] assuming that the Grüneisen parameter is only ρ-dependent. Using the EoS [80], all the thermodynamic quantities become functions of depth only, so that the reference profiles can be obtained by quadratures.

7.02.4.3.2 Perturbations of the hydrostatic, adiabatic solution

We can now rewrite the equations of fluid dynamics in terms of perturbations to the reference hydrostatic and adiabatic state (see also Alboussiere and Ricard, 2013; Bercovici et al., 1992; Glatzmaier, 1988; Jarvis and Mckenzie, 1980). The exercise is not easy as all the thermodynamic quantities, α, K_T, C_P, and C_V, are related to pressure, temperature, and density by complex Maxwell relations. Assuming inadvertently the constancy of any thermodynamic quantities may thus lead to inconsistencies. We introduce the perturbations, noted with a prime, and assume that $T=\overline{T}+T'$, $P=\overline{P}+P'$, $\rho=\overline{\rho}+\rho'$. As the reference state is adiabatic, we simply have $\mathcal{S}=\mathcal{S}'$. The variations of density and temperature are related by the following relations:

$$\begin{aligned}\rho' &= \left(\frac{\partial\rho}{\partial S}\right)_P\mathcal{S}' + \left(\frac{\partial\rho}{\partial P}\right)_S P'\\ T' &= \left(\frac{\partial T}{\partial S}\right)_P\mathcal{S}' + \left(\frac{\partial T}{\partial P}\right)_S P'\end{aligned} \qquad [124]$$

The partial derivatives of the previously mentioned expressions can be expressed as functions of the usual thermodynamic quantities. The $\mathcal{S}$ derivatives can be deduced after some algebra starting from the Maxwell relations implied by the enthalpy $d\mathcal{H}=T\mathrm{d}\mathcal{S}+\mathrm{d}P/\rho$ and by the Gibbs potential $\mathrm{d}\mathcal{G}=-\mathcal{S}\mathrm{d}T+dP/\rho$, and using eqns [31] and [33]. The adiabatic derivatives (at constant $\mathcal{S}$) are directly obtained from the reference profiles in eqns [122] and [123]. We finally end up with two equations that we consider valid at first order:

$$\begin{aligned}\rho' &= -\frac{\overline{\alpha\rho}\,\overline{T}}{\overline{C}_P}\mathcal{S}' + \frac{\nabla\overline{\rho}\cdot\mathbf{g}}{\overline{\rho}\|\mathbf{g}\|^2}P'\\ T' &= -\frac{\overline{T}}{\overline{C}_P}\mathcal{S}' + \frac{\overline{\alpha}\overline{T}}{\overline{\rho}\overline{C}_P}P'\end{aligned} \qquad [125]$$

Taking into account the variations of density with entropy and pressure, the momentum conservation [16] becomes

$$\overline{\rho}\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = -\overline{\rho}\nabla\left(\frac{P'}{\overline{\rho}}\right) + \nabla\cdot\underline{\boldsymbol{\tau}} - \frac{\overline{\mathbf{g}\alpha\rho}\,\overline{T}}{\overline{C}_P}\mathcal{S}' \qquad [126]$$

In this equation, the pressure dependence of the density has been collected with the pressure gradient. The energy conservation can take various forms (e.g., eqn [30] or [38]), but here, we use the entropy conservation [39]:

$$\overline{\rho}\,\overline{T}\frac{\mathrm{D}\mathcal{S}'}{\mathrm{D}t} = \nabla\cdot\left(\overline{k}\nabla\left(\overline{T}+T'\right)\right) + \underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \rho H \qquad [127]$$

7.02.4.3.3 Anelastic approximation

As a principle, the validity of equations cannot depend on the units in which the quantities are expressed. The laws of physics can only relate dimensionless combinations of parameters (e.g., Barenblatt, 1996). This is fundamental in fluid dynamics where a large number of quantities appear in the equations (see also Ribe, this volume). A necessary starting point is therefore to rephrase any fluid dynamics problem involving N dimensional parameters in terms of M dimensionless quantities ($N-M\leq 0$ is the number of independent physical dimensions of the problem).

Equations [126] and [127], completed with the equation of state [125] and of mass conservation, provide a differential system that controls the behavior of a nearly adiabatic and hydrostatic fluid. One cannot perform the nondimensionalization of these equations using the variable reference profiles. We must therefore introduce constant parameters, with indexes $[\,]_0$, corresponding to some typical or mantle-averaged values of the depth-dependent reference values. We introduce, for example, $\alpha_0, K_S^0, \rho_0, \mathrm{g}_0$ or C_p^0. We can use the typical values of these parameters to estimate the different terms of the convection equations.

Without approximation, the mass conservation is

$$\frac{\partial\rho'}{\partial t} + \nabla\cdot((\overline{\rho}+\rho')\mathbf{v}) = 0 \qquad [128]$$

This equation, with the momentum and energy equations, describes not only the fluid convection but also the propagation of sound waves. These waves, with velocity close to $v_\phi=\sqrt{K_S^0/\rho_0}$, travel much faster than the typical velocity of convection, at a velocity comparable to those of the P or S waves. A typical convective velocity is given by the Stokes velocity $v_s=\rho_0\alpha_0\Delta T_{na}g_0a^2/\eta_0$ (see eqn [110]). The Mach number $M=v_s/v_\phi$, ratio of the typical fluid velocity to the sound speed (Mach, 1838–1916) is very small (with the numerical values of **Table 1**, $M=1.5\times10^{-9}$). It should therefore be safe to neglect the time derivative of the density in eqn [128]; this will only filter out the sound propagation that seems a phenomenon decoupled from mantle convection.

In the divergence term of eqn [128], we can also compare $\overline{\rho}$ and ρ'. Density, temperature, and pressure variations are related by eqn [76]:

$$\frac{\rho'}{\overline{\rho}} = -\overline{\alpha}T' + \frac{P'}{K_T} \qquad [129]$$

(a relation that can also be derived from eqn [125], by eliminating $\mathcal{S}'$ and using the relation between K_S and K_T, eqn [85]). The effect of temperature on density is small, of order $\alpha_0\Delta T_{\mathrm{na}}\approx3\times10^{-2}$ (see **Table 1**). In the convecting mantle, P' is comparable to the viscous stresses $P'\approx\eta_0 v_s/a\approx\rho_0\alpha_0\Delta T_{\mathrm{na}}g_0a$ and the density variations due to the pressure fluctuation are thus

Table 1 Typical parameter values for numerical models of mantle convection

		Mantle	*Core*	*Unit*
Size	a	3×10^{6}	3×10^{6}	m
Dynamic viscosity	η_0	10^{21}	10^{-3}	Pa s
Heat capacity	$C_P^0\approx C_V^0$	1000	700	J K^{-1} kg^{-1}
Density	ρ_0	4000	11 000	kg m^{-3}
Heat cond.	k_0	3	50	$Wm^{-1}K^{-1}$
Expansivity	α_0	2×10^{-5}	10^{-5}	K^{-1}
Temperature excess	ΔT_{na}	1500	1?	K
Radioactivity prod.	H	7×10^{-11}	0?	W kg^{-1}
Gravity	g_0	9.8	5	m s^{-2}
Incompressibility	$K_T^0\approx K_S^0$	2×10^{11}	10^{12}	Pa
Kinematic viscosity	$\nu=\eta_0/\rho_0$	2.5×10^{17}	9.1×10^{-8}	m^2 s^{-1}
Thermal diff.	$\kappa=k_0/(\rho_0C_P^0)$	7.5×10^{-7}	6.5×10^{-6}	m^2s^{-1}
	$\in$	3.0×10^{-2}	$1.0\times^{-5}$	
Dissipation number	D_0	0.59	0.21	
Grüneisen parameter	Γ	1	1.3	
Rayleigh	Ra	4.2×10^{7}	2.2×10^{27}?	
Intern. Rayleigh	Ra_H	2.4×10^{10}	0?	
Prandtl	Pr	3.3×10^{23}	1.4×10^{-2}	
Grashof/Reynolds	$Gr=Re=Ra/Pr$	1.3×10^{-16}	1.6×10^{29}?	
Mach	M	1.5×10^{-9}	$\ll1$	
	M^2Pr/Ra or $D_0\epsilon/\Gamma_0$	1.7×10^{-2}	1.6×10^{-6}	

To emphasize the drastic differences between the highly viscous mantle and a real liquid (in which shear waves do not propagate), we added estimates for the core assuming that core convection is so efficient that only 1 K of nonadiabatic temperature difference can be maintained across it. Notice that with only 1 K of temperature difference, the Rayleigh number of the fluid core would already reach 10^{27}! A Stokes velocity is not appropriate to estimate the velocity in the core or its Mach number.

$$\frac{\rho'}{\overline{\rho}}\approx\alpha_0\Delta T_{na}\frac{D_0}{\Gamma_0}\frac{C_p^0}{C_v^0}=M^2\frac{Pr}{Ra}\frac{K_s^0}{K_T^0} \qquad [130]$$

In these estimates, we used either a nondimensional dissipation number

$$D_0=\frac{\alpha_0 g_0 a}{C_p^0} \qquad [131]$$

and a Grüneisen parameter, eqn [77], or the Rayleigh number Ra defined in eqn [121] and the Prandtl, Pr, number:

$$Pr=\frac{\eta_0 C_P^0}{k_0}=\frac{\nu_0}{\kappa_0} \qquad [132]$$

The ratios C_p^0/C_v^0 and K_s^0/K_T^0 are of order 1. The dissipation number compares the natural scale of temperature variations [123] with the layer thickness, a. The dissipation number D_0 is around 0.5 for the Earth's mantle. For any laboratory experiment, this number would be infinitely small; only geophysical or astrophysical problems have large dissipation numbers. The physical meaning of the Rayleigh number as a measure of convective vigor has already been discussed. The Prandtl number (Prandtl, 1975–53) compares the two diffusive processes: namely, the diffusion of momentum and heat. As $D_0\approx0.5$ and $\Gamma_0\approx1$, the density variations due to non hydrostatic pressure variations are smaller but of the same order as those due to temperature variations. There are both of a few percent so that $\rho'\ll\overline{\rho}$. The mass conservation can therefore be approximated in what is called the anelastic approximation (Ogura and Phillips, 1962; Spiegel and Veronis, 1960):

$$\nabla\cdot(\overline{\rho}\mathbf{v})=0 \qquad [133]$$

The anelastic approximation is sometimes called small Mach number approximation. However, the anelastic approximation really requires not only a small M but also a small M^2Pr/Ra (Bercovici et al., 1992). A planet could have a very low Mach number but so large a Prandtl number that the anelastic approximation would not be valid. Similarly, a planet can have a low Reynolds number (creeping convection) with a very large Rayleigh number (chaotic convection).

The differential system eqns [126], [127], [125], and [133] constitute the anelastic approximation of convection that has been used for atmospheric convection as well as for liquid and solid convection, including mantle and core convection. A large difference between gases and liquids or solids is that the former have $\alpha T\approx1$ (and exactly 1 for ideal gas) and the latter have $\alpha T\ll1$. This allows a further simplification of the equation, the anelastic liquid approximation.

7.02.4.3.4 Anelastic liquid approximation

Using the relation between entropy, pressure, and temperature (eqn [125]), and following the same scaling as before, where non hydrostatic pressure and viscous stresses have the same magnitude, yields

$$T'=\frac{\overline{T}}{\overline{C}_P}\mathcal{S}'+\frac{\overline{\alpha}\overline{T}}{\overline{\rho}\overline{C}_P}P'\approx\frac{\overline{T}}{\overline{C}_P}S'+D_0\overline{\alpha}\overline{T}T' \qquad [134]$$

As $D_0\approx0.5$ and $\overline{\alpha}\overline{T}\ll0$, the last term is negligible, and for the mantle or the core, we can identify $\overline{T}\mathcal{S}'$ and $\overline{C}_PT'$. Using $\overline{T}\mathrm{d}\mathcal{S}'=\mathrm{d}(\overline{T}\mathcal{S}')-\mathcal{S}'\mathrm{d}\overline{T}=\mathrm{d}(\overline{C}_PT')-\overline{C}_P(T'/\overline{T})\mathrm{d}\overline{T}$, we can express the anelastic liquid equations by

$$\begin{aligned}&\nabla\cdot(\rho'\mathbf{v})=0\\&\overline{\rho}\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t}=-\overline{\rho}\nabla\left(\frac{P'}{\overline{\rho}}\right)+\nabla\cdot\underline{\boldsymbol{\tau}}-\overline{\rho}\,\overline{\mathbf{g}}\overline{\alpha}T'\\&\overline{\rho}\frac{\mathrm{D}\overline{C}T'}{\mathrm{D}t}=\overline{\rho}\,\overline{\mathrm{g}}\overline{\alpha}v_gT'+\nabla\cdot\left(\overline{k}\nabla\left(\overline{T}+T'\right)\right)+\underline{\boldsymbol{\tau}}:\nabla\mathbf{v}+\rho H\end{aligned} \qquad [135]$$

where v_g is the velocity along $\overline{\mathrm{g}}$.

In the anelastic liquid approximation, C_P and C_V are equal, as well as K_S and K_T (their relative difference is $\Gamma_0\alpha_0\Delta T_{na}\ll1$

according to eqn [85]); we thus simply label the heat capacity and incompressibility as $\overline{C}$ and $\overline{K}$.

7.02.4.3.5 Nondimensionalization

To nondimensionalize the equations, we perform the change of variables and parameters

$$\begin{aligned} &x,\ v,\ t,\ P',\ T',\ \tau \rightarrow a\widetilde{x}, v_s\widetilde{v},\ \frac{a}{v_s}\widetilde{t},\ \frac{\eta_0 v_s}{a}\widetilde{P}',\ \Delta T_{na}\widetilde{T}',\ \frac{\eta_0 v_S}{a}\widetilde{\tau} \\ &\nabla,\ \overline{\rho},\ \overline{\alpha},\ \overline{C},\ \overline{K},\ \overline{k},\ \overline{g} \rightarrow \frac{1}{a}\widetilde{\nabla},\ \rho_0\widetilde{\rho},\ \alpha_0\widetilde{\alpha},\ C_0\widetilde{C},\ K_0\widetilde{K},\ k_0\widetilde{k},\ g_0\widetilde{g} \end{aligned} \quad [136]$$

where v_s is the Stokes velocity $a^2 g_0\rho_0\alpha_0\ \Delta T_{na}/\eta_0$. We obtain the nondimensionalized fluid anelastic equations:

$$\begin{aligned} &\nabla\cdot(\widetilde{\rho}\mathbf{v}) = 0 \\ &\widetilde{\rho}\frac{Ra}{Pr}\frac{D\mathbf{v}}{Dt} = -\widetilde{\rho}\nabla\left(\frac{P}{\widetilde{\rho}}\right) + \nabla\cdot\underline{\tau} - \widetilde{\rho}\widetilde{\mathbf{g}}\widetilde{\alpha}T \\ &\widetilde{\rho}\frac{D\widetilde{C}\widetilde{T}}{Dt} = \widetilde{\rho}\widetilde{g}\widetilde{\alpha}v_g D_0 T + \frac{1}{Ra}\nabla\cdot\left(\widetilde{k}\nabla\left(\widetilde{T}+T\right)\right) \\ &\qquad + D_0\underline{\tau}:\nabla\mathbf{v} + \frac{1}{Ra}\widetilde{\rho}\frac{\rho_0 Ha^2}{k_0\Delta T_{na}} \end{aligned} \quad [137]$$

For simplicity, we only kept the tilde sign on the reference nondimensionalized profiles, and we omitted the tilde and the primes on all the other quantities. In the momentum equation, the nondimensionalized stress tensor in the Newtonian case (without bulk viscosity) is

$$\underline{\tau} = \widetilde{\eta}\left(\nabla\mathbf{v} + [\nabla\mathbf{v}]^t\right) - \frac{2}{3}\widetilde{\eta}\nabla\cdot\mathbf{v} \quad [138]$$

The formalism is already so heavy that we have not included the self-gravitational term. This term is not negligible at long wavelengths (see Section 7.02.2.5.2). To account for this term, we should have added on the right side of momentum equation of [137] a term $-\widetilde{\rho}\nabla\widetilde{\psi}$. where the perturbed gravitational potential due to the departure of the density from the reference profile satisfies Poisson's equation:

$$\nabla^2\widetilde{\psi} = 3\widetilde{\rho}\frac{\rho_0}{<\rho>}\left(\frac{1}{\widetilde{K}}\frac{D_0}{\Gamma_0}P' - \widetilde{\alpha}T'\right) \quad [139]$$

where $<\rho>$ is the average density of the Earth.

7.02.4.3.6 Dimensionless numbers

The ratio Ra/Pr is also called the Grashof number $Gr = \alpha_0 g_0 \Delta T_{na} a^3/\nu_0^2$ (Grashof, 1826–93). Introducing the Stokes velocity v_S, eqn [110], this number can also be written as $v_S a/\nu_0$ and could be called the Reynolds number, Re, of the flow (Reynolds, 1842–1912). Using one or the other names depends on the quantities that are best known. For example, if the velocity V_0 is a parameter imposed by a boundary condition, using it to perform the nondimensionalization and characterizing the system by the Reynolds number $V_0 a/\nu_0$ would be more natural than using the Grashof number. If a thermal structure is imposed by a velocity boundary condition (e.g., by the thickening of the oceanic lithosphere with age), it would seem natural to introduce a Péclet number $V_0 a/\kappa_0$ (Péclet, 1793–1857), which is nothing more than the Rayleigh number of the flow if the velocity is imposed by the convection.

The Rayleigh and Prandtl numbers can be estimated in different ways. In fact, the only parameter difficult to know is the viscosity. In most textbooks, the value of 10^{21} Pa s, first proposed by Haskell (1937), is given with a unanimity that hides very large uncertainties and most probably a large geographic variability. The mantle viscosity and its depth dependence can be constrained by postglacial rebound, geoid, true polar wander, change of flattening of the Earth, and plate force balance models or extrapolated from laboratory measurements. An increase of viscosity with depth, either one or two orders of magnitudes, is likely, with an asthenosphere significantly less viscous (10^{19} Pa s) at least under oceanic plates and a lower mantle probably around 10^{22} Pa s (see details in Section 7.02.6.1). It is impossible to give justice to all the papers on this subject, but some geodynamic estimates of mantle viscosity can be found in, for example, Peltier (1989), Sabadini and Yuen (1989), Lambeck and Johnston (1998), and Ricard et al. (1993a). Whatever the value of the real viscosity, the ratio Ra/Pr is so small that inertia plays no role in the mantle and the left side of the momentum equation in the anelastic approximation [137] can safely be set to zero (see numerical values in Table 1).

7.02.4.3.7 Boussinesq approximation

As was expected from Section 7.02.2.5.3, where we had shown that the dissipation and the adiabatic terms balance each other in a statistical steady-state regime, these terms are proportional to the same dissipation number D_0. Although D_0 is not so small, most of the physics of mantle convection, except for the additional adiabatic temperature gradient, is captured with models where D_0 is arbitrarily set to zero. The Boussinesq approximation captures most of the mantle dynamics due to the fact that a huge ΔT_{na} is maintained across the top and bottom boundary layers. The other sources and sinks of energy and density anomaly (viscous dissipation and adiabatic heating) are therefore always small (Bercovici et al., 1992; Glatzmaier, 1988; Jarvis and Mckenzie, 1980; Tackley, 1996).

In the Boussinesq approximation, $D_0 = 0$, the reference density and temperature as well as the heat capacity and the thermal expansivity become constants (according to eqn [123], $\nabla\widetilde{T} \approx D_0\widetilde{T}$ and $\nabla\widetilde{\rho} \approx (D_0/\Gamma_0)\widetilde{\rho}$)). The nonadiabatic temperature increase ΔT_{na} becomes simply the total temperature increase ΔT. This approximation is of course excellent for laboratory-scale experiments where effectively $D_0 \ll 1$. The EoS [130] indicates that the density variations due to pressure are D_0/Γ_0 smaller than those due to temperature and are therefore negligible and the fluid dynamics equations become

$$\begin{aligned} &\nabla\cdot\mathbf{v} = 0 \\ &\frac{Ra}{Pr}\frac{D\mathbf{v}}{Dt} = -\nabla P + \nabla\cdot\left(\widetilde{\eta}\left(\nabla\mathbf{v} + [\nabla\mathbf{v}]^t\right)\right) - \widetilde{\mathbf{g}}T \\ &\frac{DT}{Dt} = \frac{1}{Ra}\nabla\cdot\left[\widetilde{k}\nabla T\right] + \frac{1}{Ra}\frac{\rho_0 Ha^2}{k_0\Delta T} \end{aligned} \quad [140]$$

Here again, as in the mantle, $Gr = Ra/Pr \ll 1$, the inertia in the momentum equation [140] can be neglected. The self-gravitational term $-\widetilde{\rho}\widetilde{\nabla}\widetilde{\psi}$ should be added to the momentum equation (second equation of eqn [140]) for large-scale simulations, the gravitational potential being solution of eqn [139] where only the thermal part of the density variation needs to be taken into account.

The nondimensionalized gravity $\widetilde{\mathbf{g}}$, viscosity $\widetilde{\eta}$, and thermal conductivity $\widetilde{k}$ are often assumed constant (e.g., $\widetilde{k} = \widetilde{\eta} = 1$ and $\widetilde{\mathbf{g}} = \widetilde{\mathbf{e}}_g$, a constant unit vector), but they can be variable. The

physical behavior of a large Rayleigh number is obvious in eqn [140]. When $Ra \to \infty$, and in the absence of heat sources, the temperature becomes a purely advected and conserved quantity, $\mathrm{D}T/\mathrm{D}t = 0$.

7.02.4.3.8 Internal heating

In the nondimensionalization, we assumed that the nonadiabatic temperature ΔT_{na} and the radioactive sources are two independent quantities. Of course, in the case where the mantle is only heated from within, the excess temperature is not a free parameter but must result from the heat source and from the flow itself. In the nondimensionalization, we can replace ΔT_{na} by $\rho_0\, Ha^2/k_0$ in such a way that the radioactive heat source of the anelastic or Boussinesq energy equations, [137] and [140], are simply $1/Ra$. This choice requires the introduction of a somewhat different Rayleigh number, the internally heated Rayleigh number (Roberts, 1967),

$$Ra_{\mathrm{H}} = \frac{\alpha_0 H \rho_0^3 g_0 a^5 C_P^0}{\eta_0 k_0^2} \qquad [141]$$

7.02.4.3.9 Change of nondimensionalization

We use a Stokes velocity to nondimensionalize the equations. We have therefore introduced a velocity V_0 of order of 300 m year^{-1} and a time $a/V_0 = 10\,000$ year (see **Table 1**). This is certainly very fast and short compared with geologic scales. Most physical and geophysical textbooks (e.g., Schubert et al., 2001) use instead a diffusive time $t_{\mathrm{D}} = \rho C_P a^2/k_0$ and velocity a/t_{D}. This is perfectly valid, but **Table 1** shows that the diffusive time and velocity amount to $t_{\mathrm{D}} = 400$ byr and $V_{\mathrm{D}} = 7 \times 10^{-6}$ m year^{-1}. These values would on the contrary be very slow and long compared with geologic scales. A nondimensionalization using a diffusive timescale leads to the anelastic equations

$$\begin{aligned}
&\nabla\cdot(\widetilde{\rho}\mathbf{v}) = 0\\
&\widetilde{\rho}\frac{1}{Pr}\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = -\widetilde{\rho}\nabla\left(\frac{P}{\widetilde{\rho}}\right) + \nabla\cdot\underline{\boldsymbol{\tau}} - Ra\widetilde{\rho}\widetilde{\mathbf{g}}\widetilde{\alpha}T\\
&\widetilde{\rho}\frac{\mathrm{D}\widetilde{C}T}{\mathrm{D}t} = \widetilde{\rho}\widetilde{\mathbf{g}}\widetilde{\alpha}D_0\upsilon_{\mathrm{g}}T + \nabla\cdot\left(\widetilde{k}\nabla\left(\widetilde{T}+T\right)\right) + \frac{D_0}{Ra}\underline{\boldsymbol{\tau}}:\nabla\mathbf{v} + \frac{\rho_0 Ha^2}{k_0\Delta T_{\mathrm{na}}}
\end{aligned} \qquad [142]$$

and the Boussinesq equations

$$\begin{aligned}
&\nabla\cdot(\mathbf{v}) = 0\\
&\frac{1}{Pr}\frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = -\nabla P + \nabla\cdot\underline{\boldsymbol{\tau}} - Ra\widetilde{\mathbf{g}}T\\
&\frac{\mathrm{D}T}{\mathrm{D}t} = \nabla\cdot\left(\widetilde{k}\nabla T\right) + \frac{\rho_0 Ha^2}{k_0\Delta T}
\end{aligned} \qquad [143]$$

Notice that the Ra number appears in different places than in eqns [137] and [140]. Of course, after their appropriate changes of variables, the dimensional solutions are the same.

7.02.4.4 Linear Stability Analysis for Basally Heated Convection

To understand why the static conductive solution is not necessarily the solution chosen by the system, one can test the stability of a static state. This is what physicists call a study of stability analysis (see also Ribe, this volume). It consists of substituting into the basic equations the known static solution plus an infinitely small perturbation and checking whether or not this perturbation amplifies, decreases, or propagates. It is only when perturbations decrease in amplitude that the tested equilibrium is established to be stable.

To illustrate this analysis, we use the Boussinesq approximation, with constant viscosity and conductivity, neglecting inertia and without internal heating. The nondimensionalized equations [140] yield

$$\begin{aligned}
\nabla\cdot\mathbf{v} &= 0\\
-\nabla P + \nabla^2\mathbf{v} - T\mathbf{e}_z &= 0\\
\frac{\partial T}{\partial t} + \mathbf{v}\cdot\nabla T &= \frac{1}{Ra}\nabla^2 T
\end{aligned} \qquad [144]$$

($\mathbf{e}_z$ is the unit vector parallel to $\mathbf{g}$). The steady diffusive nondimensional temperature solution is $T = z$, and we test a solution of the form $T = z + \delta T$. The temperature boundary condition, $T = 0$ on top and $T = 1$ at the bottom, requires that δT vanishes for $z = 0$ and $z = 1$. In static conductive state, the velocity is zero, thus the velocity $\mathbf{v}$ induced by the perturbation δT will be infinitesimally small. In the nonlinear term, we can approximate $\mathbf{v}\cdot\nabla T = \mathbf{v}\cdot\nabla(z + \delta T) \approx \upsilon_z$. With this approximation, the equations are linear and we can find a solution in the form of a plane wave.

For a fluid confined between $z = 0$ and $z = 1$ where the temperature is imposed (i.e., where $\delta T = 0$) and unbounded in the x-direction, a solution, $\delta T = \theta(t)\sin(\pi z)\sin(kx)$, is appropriate and satisfies the boundary conditions. This solution is 2-D, has a single mode in the z-direction, and is periodic in x with wavelength $\lambda = 2\pi/k$. More complex patterns could be tried but the mode we have chosen would destabilize first (see Ribe, this volume). It is then straightforward to deduce that for such a thermal anomaly, the energy equation imposes a vertical velocity

$$\upsilon_z = -\left(\dot{\theta} + \frac{(k^2+\pi^2)}{Ra}\theta\right)\sin(\pi z)\sin(kx) \qquad [145]$$

From mass conservation and assuming top and bottom free-slip conditions, the x-component of the velocity must be

$$\upsilon_x = -\frac{\pi}{k}\left(\dot{\theta} + \frac{(k^2+\pi^2)}{Ra}\theta\right)\cos(\pi z)\cos(kx) \qquad [146]$$

When the velocity and the temperature are introduced in the momentum equations, the pressure eliminated, the time evolution of the temperature perturbation is found

$$\dot{\theta} = \theta\left(\frac{k^2}{(\pi^2+k^2)^2} - \frac{(\pi^2+k^2)}{Ra}\right) \qquad [147]$$

For any wave number k, a small enough Rayleigh number corresponds to a stable solution, $\dot{\theta}/\theta < 0$. When the Rayleigh number is increased, the temperature perturbation with wave number k becomes unstable at the threshold Rayleigh number

$$Ra = \frac{(\pi^2+k^2)^3}{k^2} \qquad [148]$$

This $Ra(k)$ curve is plotted in **Figure 3**. This curve has a minimum when

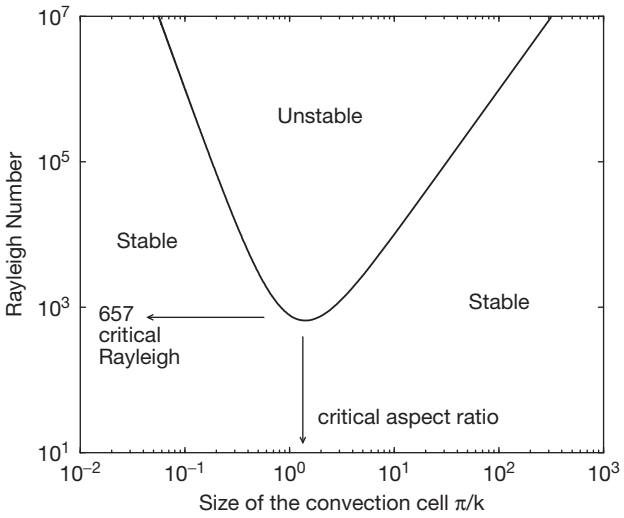


Figure 3 Critical Rayleigh number as a function of the half wavelength π/k (the size of the convection cells). Above this curve, convection occurs with a whole range of unstable wavelengths. Below this curve, the conductive temperature is stable since temperature perturbations of any wavelength decrease. When the Rayleigh number is increased, the first unstable wavelength corresponds to a convection cell of aspect ratio $\sqrt{2}$ and a critical Rayleigh number of 657.

$$k=\frac{\pi}{\sqrt{2}},\quad Ra_c=\frac{27}{4}\pi^4\approx 657 \qquad [149]$$

What can be interpreted as the size of one convective cell is π/k since one wavelength corresponds to two counterrotating cells. The critical cell has an aspect ratio, width over height, of $\sqrt{2}$.

A Rayleigh number of 657 is the critical Rayleigh number for the onset of convection in a layer heated from below with free-slip boundary conditions. As soon as $Ra > Ra_c$, there is a wave number interval over which convection begins. Of course, when convection grows in amplitude, the marginal stability solution becomes less and less pertinent as the assumption that $\mathbf{v}\cdot\boldsymbol{\nabla}\delta T \ll \mathbf{v}\cdot\boldsymbol{\nabla}z$ becomes invalid. The existence of a critical Rayleigh number and its value, here obtained from a marginal stability analysis, have also been obtained in a more general case, where finite amplitude perturbations (instead of infinitesimal ones) are added to the stationary solution. Below the critical Rayleigh number, any finite perturbation can be shown to decay eventually to zero (e.g., Joseph, 1966).

7.02.4.5 Road to Chaos

In Cartesian geometry, when the Rayleigh number reaches its critical value, convection starts and forms rolls. When the Rayleigh number is further increased, complex series of convection patterns can be obtained, first stationary, then periodic, and finally chaotic (see **Chapter 7.03**). Using the values of **Table 1**, the critical Rayleigh number of the mantle would be attained for a nonadiabatic temperature difference between the surface and the CMB of only 0.025 K. The mantle Rayleigh number is several orders of magnitude higher than critical and the mantle is in a chaotic state of convection.

Figure 4 shows a stationary convection pattern at $Ra=10^5$ and three snapshots of numerical simulation of convection at higher Rayleigh number. The color scale has been chosen differently in each panel to emphasize the thermal structures that decrease in length scale with Ra. This view is somewhat misleading since all the thermal anomalies become confined in a top cold boundary layer and in a hot bottom one at large Rayleigh numbers. Most of the interior of the cell becomes just isothermal (or adiabatic when anelastic equations are used). The various transitions of convection as the Rayleigh number increases will be discussed in other chapters of this treatise (see **Chapters 7.03**, **7.04**, **7.05**).

7.02.5 Introduction to Physics of Multicomponent and Multiphase Flows

The mantle is not a simple homogeneous material. It is made of grains of variable bulk composition and mineralogy and contains fluids, magma, and dissolved volatiles. Discussion of multicomponent and multiphase flows could deal with solids, liquids, or gases; include compressibility or not; and consider elastic, viscous, or more complex rheology. For each combination of these characteristics, a geophysical application is possible. Here, we will restrict the presentation to viscous creep models (i.e., without inertia), where the various components are treated with continuous variables (i.e., each component is implicitly present everywhere). We do not consider approaches where the various components are separated by moving and deformable interfaces. Our presentation excludes cases where the problem is to match properties at macroscopic interfaces between regions of different but homogeneous compositions.

We will focus on two cases. The first will be when all the components are perfectly mixed in variable proportions. This corresponds to the classical chemical approach of multiple components in a solution. This will provide some tools to understand mantle phase transitions and the physics of chemical diffusion and mixing. We will be rather formal and refer to other chapters of this treatise for the applications and illustrations (e.g., Parmentier and Tackley, this volume). Our goal is to explain why and when the advection–diffusion equation can be used in mantle dynamics. The irreversible thermodynamics of multicomponent flows is discussed in various classical books (e.g., de Groot and Mazur, 1984; Haase, 1990). However, as usual with geophysical flows, the mantle has many simplifications and a few complexities that are not necessarily well documented in these classical textbooks.

The second case will be for two-phase flows in which the two phases are separated by physical interfaces that are highly convolved and with spatial characteristics much smaller than the typical size of geodynamic models. This is typically the case where magma can percolate through a compacting matrix (see also Dingwell, vol. 2). This approach was used to model melt extraction and core–mantle interaction (McKenzie, 1984; Scott and Stevenson, 1984). Magma migration has also been treated in a large number of publications where solid and magma are considered as separated in studies of dike propagation through hydraulic fracturing (e.g., Lister and Kerr, 1991) or where fusion is parameterized in some way (e.g., Choblet and Parmentier, 2001; Ito et al., 1999). We do not discuss these latter approaches.

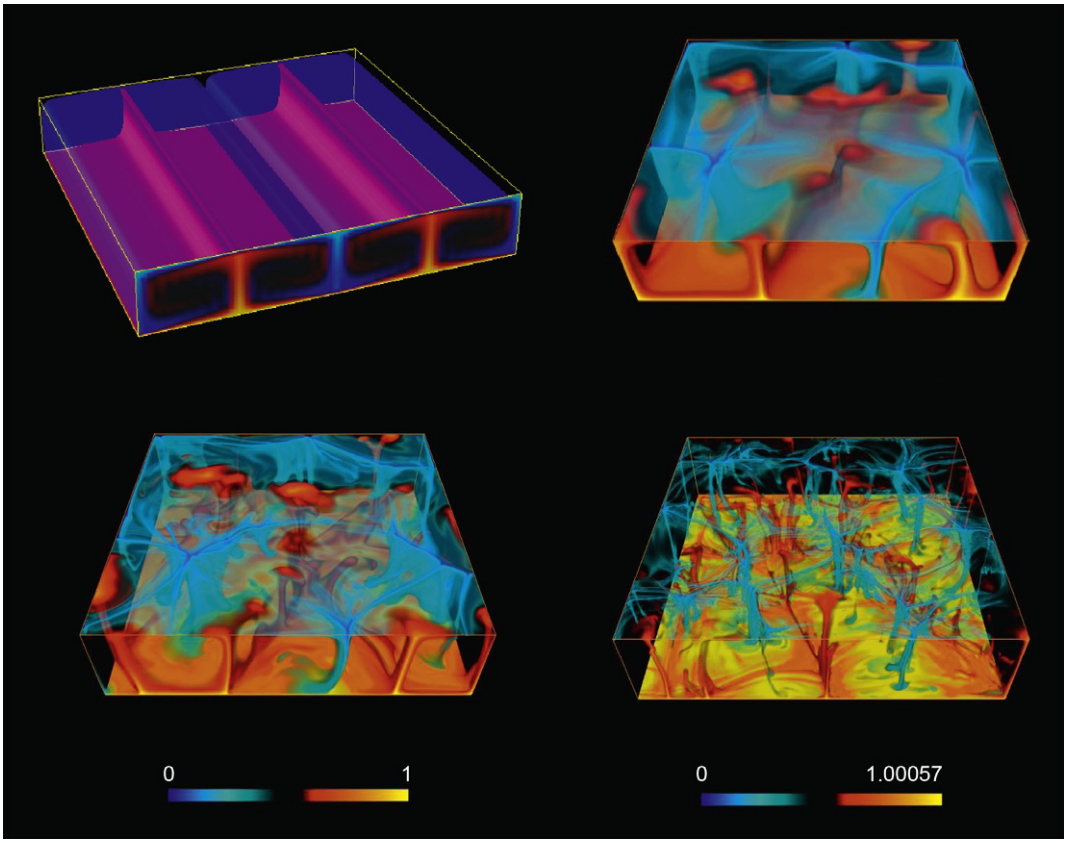


Figure 4 Convection patterns of a fluid heated from below at Rayleigh number 10^5, 10^6, 10^7, and 10^8. The temperature color bars range from 0 (top boundary) to 1 (bottom boundary). The Boussinesq approximation was used (numerical simulations by F. Dubuffet). The increase in Rayleigh number corresponds to a decrease of the boundary layer thicknesses and the width of plumes. Only in the case of the lowest Rayleigh number (top left) is the convection stationary with cells of aspect ratio $\approx \sqrt{2}$ as predicted by marginal stability. For higher Rayleigh number, the patterns are highly time-dependent.

7.02.5.1 Fluid Dynamics of Multicomponent Flows in Solution

7.02.5.1.1 Mass conservation in a multicomponent solution

If we want to study the evolution of major or trace element concentration in the convecting mantle, we can consider the mantle, instead of a homogeneous fluid, as a solution of various components i in volumetric proportions ϕ_i (with $\sum_i \phi_i = 1$) having the densities ρ_i and velocities $\mathbf{v}_i$ (and later, thermal expansivities α_i and heat capacities $C_p^i \ldots$).

Using a mass balance very similar to what we had discussed for a homogeneous fluid, we obtain a mass conservation equation of the form

$$\frac{\partial(\phi_i \rho_i)}{\partial t} + \boldsymbol{\nabla}\cdot(\phi_i \rho_i \mathbf{v}_i) = \Gamma_i \qquad [150]$$

where Γ_i is the rate of mass production of component i. This rate of mass production is zero if no reactions produce the component i.

In the fluid, the average density is

$$\overline{\rho} = \sum_i \phi_i \rho_i \qquad [151]$$

and various average velocities can be defined (weighted by the mass, the volume, the number of moles, etc., of each component i). In this section, we introduce the barycentric velocity, $\mathbf{v}_\mathrm{b}$ (velocity of the center of mass), defined by

$$\mathbf{v}_\mathrm{b} = \frac{\sum \phi_i \rho_i \mathbf{v}_i}{\overline{\rho}} \qquad [152]$$

The average mass conservation can be obtained by summing the equations of component conservation [150],

$$\frac{\partial \overline{\rho}}{\partial t} + \nabla\cdot(\overline{\rho}\mathbf{v}_\mathrm{b}) = 0 \qquad [153]$$

since the sum of the rates of mass production is zero:

$$\sum_i \Gamma_i = 0 \qquad [154]$$

In equation [150], instead of the various component velocities $\mathbf{v}_i$, we can introduce the barycentric velocity $\mathbf{v}_\mathrm{b}$ and the diffusive flux of the component i with respect to this average flow,

$$\frac{\partial(\phi_i \rho_i)}{\partial t} + \boldsymbol{\nabla}\cdot(\phi_i \rho_i \mathbf{v}_\mathrm{b}) = -\boldsymbol{\nabla}\cdot\mathbf{J}_i + \Gamma_i \qquad [155]$$

where we define the diffusive flux, $\mathbf{J}_i$, by

$$\mathbf{J}_i = \phi_i \rho_i (\mathbf{v}_i - \mathbf{v}_\mathrm{b}) \qquad [156]$$

By definition of the barycentric velocity eqn [152], the diffusive flows sum to zero:

$$\sum_i \mathbf{J}_i = 0 \qquad [157]$$

Diffusive transport of mass is nothing more than the advective transfer by component velocities relative to the average barycentric velocity. We will show later in simple cases that the diffusive fluxes are driven by concentration gradients (de Groot and Mazur, 1984; Haase, 1990; Woods, 1975).

If we introduce the mass fraction $C_i = \phi_i \rho_i / \overline{\rho}$ (in kg of i per kg of mixture), we can easily show from eqns [153] and [155] that

$$\overline{\rho} \frac{\mathrm{D}C_i}{\mathrm{D}t} = -\boldsymbol{\nabla}\cdot\mathbf{J}_i + \Gamma_i \qquad [158]$$

where the Lagrangian derivative is defined with the barycentric velocity,

$$\frac{\mathrm{D}}{\mathrm{D}t} = \frac{\partial}{\partial t} + \mathbf{v}_\mathrm{b}\cdot\boldsymbol{\nabla} \qquad [159]$$

7.02.5.1.2 Momentum and energy in a multicomponent solution

In our multicomponent solution, all constituents are present at each point, and they are all locally submitted to the same pressure and stresses. We assume that the viscous stress is simply related to $\mathbf{v}_\mathrm{b}$ and we neglect inertia as appropriate for the mantle. Newton's second law (here, simply the balance of forces) can be applied to the barycenter and implies

$$\boldsymbol{\nabla}\cdot\underline{\boldsymbol{\tau}} - \boldsymbol{\nabla}P + \overline{\rho}\mathbf{g} = 0 \qquad [160]$$

where the only force is due to the (constant) gravity. The momentum equation thus remains identical to that of a fluid with uniform composition and without inertia (eqn [16]).

Since there is only one momentum equation for i components, the $i-1$ other velocity equations will be found by using the constraints of the laws of thermodynamics and in particular the positivity of the entropy source. To derive the energy conservation, we perform the standard balance to account for all the energy exchanges in a volume Ω and across its surface $\sum$. Instead of the one component equation [30], we have to sum up various contributions and we get

$$\sum_i \frac{\partial(\phi_i\rho_i\mathcal{U}_i)}{\partial t} = -\boldsymbol{\nabla}\cdot\left(\sum_i \phi_i\rho_i\mathcal{U}_i\mathbf{v}_i + P\sum_i \phi_i\mathbf{v}_i + \mathbf{q} - \underline{\boldsymbol{\tau}}\cdot\mathbf{v}_\mathrm{b}\right) + \mathbf{g}\cdot\sum_i \phi_i\rho_i\mathbf{v}_i + \overline{\rho}\overline{H} \qquad [161]$$

In this expression, we recognize the temporal changes in energy ($\mathcal{U}_i$ is the internal energy per unit mass of component i and the kinetic energies are neglected), the bulk energy flux, the pressure work, the thermal diffusion, the viscous stress work, the gravity work, and the radioactivity production $\left(\overline{\rho}\overline{H} = \sum_i \phi_i\rho_i H_i\right)$. The various ϕ_i come from the assumption that each component i is present in proportion ϕ_i in both the volume Ω and its surface $\sum$. We assume that thermal diffusion acts equally for each component and that the surface work of the stress tensor is only related to the barycentric velocity.

Using the definition of the barycentric velocity [152], of the diffusive fluxes [156], and of the momentum conservation [160] and using $\sum_i \phi_i = 1$, the energy expression can be simplified to

$$\sum_i \phi_i\rho_i \frac{\mathrm{D}\mathcal{H}_i}{\mathrm{D}t} = -\boldsymbol{\nabla}\cdot\mathbf{q} - \sum_i \mathbf{J}_i\cdot\boldsymbol{\nabla}\mathcal{H}_i + \frac{\mathrm{D}P}{\mathrm{D}t} - \sum_i \Gamma_i\mathcal{H}_i + \underline{\boldsymbol{\tau}} : \nabla\mathbf{v}_\mathrm{b} + \overline{\rho}\overline{H} \qquad [162]$$

where $\mathcal{H}_i$ are the component enthalpies

$$\mathcal{H}_i = \mathcal{U}_i + \frac{P}{\rho_i} \qquad [163]$$

The enthalpy variation for each component i can be expressed as a function of the state variables P and T. According to eqn [60], we can write for each component

$$\rho_i \frac{\mathrm{D}\mathcal{H}_i}{\mathrm{D}t} = \rho_i C_p^i \frac{\mathrm{D}T}{\mathrm{D}t} + (1 - \alpha_i T)\frac{\mathrm{D}P}{\mathrm{D}t} \qquad [164]$$

Finally, the expression for the temperature evolution is

$$\overline{\rho}\overline{C}_p \frac{\mathrm{D}T}{\mathrm{D}t} = -\boldsymbol{\nabla}\cdot\mathbf{q} - \sum_i \mathbf{J}_i\cdot\boldsymbol{\nabla}\mathcal{H}_i + \overline{\alpha}T\frac{\mathrm{D}P}{\mathrm{D}t} - \sum_i \Gamma_i\mathcal{H}_i + \underline{\boldsymbol{\tau}} : \boldsymbol{\nabla}\mathbf{v}_\mathrm{b} + \overline{\rho}\overline{H} \qquad [165]$$

where the average heat capacity and thermal expansivity are $\overline{C}_p = \sum_i \phi_i\rho_i C_p^i / \overline{\rho}$ and $\overline{\alpha} = \sum_i \phi_i \alpha_i$.

Compared to the homogeneous case [38], two new heat source terms are present, the enthalpy exchange through chemical reactions, $\sum_i \Gamma_i\mathcal{H}_i$, and the enthalpy redistribution by component diffusion, $\sum_i \mathbf{J}_i\cdot\boldsymbol{\nabla}\mathcal{H}_i$.

7.02.5.1.3 Entropy conservation in a multicomponent solution

Entropy conservation is essential for deriving the expressions of the diffusive fluxes. The general expression of entropy conservation [7] is

$$\sum_i \frac{\partial(\phi_i\rho_i\mathcal{S}_i)}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{J}_\mathcal{S} + H_\mathcal{S} \qquad [166]$$

where $\mathbf{J}_\mathcal{S}$ and $H_\mathcal{S}$ are the yet unknown entropy flux and source. The entropy of the various components takes into account their specific entropies as well as their configurational entropies or mixing entropies due to the dispersion of the component i in the solution. Introducing the barycentric velocities and the diffusive fluxes, this equation can be recast as

$$\sum_i \phi_i\rho_i \frac{\mathrm{D}\mathcal{S}_i}{\mathrm{D}t} = \boldsymbol{\nabla}\cdot\left(\sum_i \phi_i\rho_i\mathcal{S}_i\mathbf{v}_\mathrm{b} + \sum_i \mathcal{S}_i\mathbf{J}_i - \mathbf{J}_\mathcal{S}\right) + H_\mathcal{S} - \mathcal{S}_i\Gamma_i - \mathbf{J}_i\cdot\boldsymbol{\nabla}\mathcal{S}_i \qquad [167]$$

However, a second expression of the entropy conservation can be obtained from the enthalpy conservation, [162]: using $\mathrm{d}\mathcal{H}_i = T\mathrm{d}\mathcal{S}_i + V_i\mathrm{d}P$, which, in our case, can be expressed as

$$\rho_i \frac{\mathrm{D}\mathcal{H}_i}{\mathrm{D}t} = \rho_i T\frac{\mathrm{D}\mathcal{S}_i}{\mathrm{D}t} + \frac{\mathrm{D}P}{\mathrm{D}t} \qquad [168]$$

we derive

$$\sum_i \phi_i\rho_i T\frac{\mathrm{D}\mathcal{S}_i}{\mathrm{D}t} = -\boldsymbol{\nabla}\cdot\mathbf{q} - \sum_i \mathbf{J}_i\cdot\boldsymbol{\nabla}\mathcal{H}_i - \sum_i \Gamma_i\mathcal{H}_i + \underline{\boldsymbol{\tau}} : \boldsymbol{\nabla}\mathbf{v}_\mathrm{b} + \overline{\rho}\overline{H} \qquad [169]$$

A comparison of the two expressions for the entropy conservation, [167] and [169], allows us to identify the total entropy flux

$$\mathbf{J}_S = \frac{\mathbf{q}}{T} + \mathbf{v}_b \sum_i \phi_i \rho_i \mathcal{S}_i + \sum_i \mathcal{S}_i \mathbf{J}_i \qquad [170]$$

and the entropy sources

$$TH_S = -\left(\frac{\mathbf{q}}{T} + \sum_i \mathcal{S}_i \mathbf{J}_i\right) \cdot \nabla T - \sum_i \mathbf{J}_i \cdot \nabla \mu_i - \sum_i \Gamma_i \mu_i + \underline{\boldsymbol{\tau}} : \nabla \mathbf{v}_b + \overline{\rho} \overline{H} \qquad [171]$$

where we introduced the chemical potentials $\mu_i = \mathcal{H}_i - T\mathcal{S}_i$. The total entropy flux, eqn [170], is related to thermal diffusion and to advection and chemical diffusion of component entropies.

In eqn [171], the gradients of chemical potential and temperature are not independent as the chemical potential gradients implicitly include the temperature gradient, so that alternative expressions can be found. For example, we can introduce the gradient of μ at constant temperature $\nabla_T \mu$ as

$$\nabla_T \mu_i = \nabla \mu_i + \mathcal{S}_i \nabla T \qquad [172]$$

which leads to

$$TH_S = -\frac{1}{T}\mathbf{q} \cdot \nabla T - \sum_i \mathbf{J}_i \cdot \nabla_T \mu_i - \sum_i \Gamma_i \mu_i + \underline{\boldsymbol{\tau}} : \nabla \mathbf{v}_b + \overline{\rho} \overline{H} \qquad [173]$$

This last equation has the advantage of separating the temperature contribution, ∇T, from the compositional contribution, $\nabla_T \mu_i$ ($\nabla_T \mu_i$ varies mostly with composition as composition can change over very short distances; however, this term is also related to pressure variations) (see de Groot and Mazur, 1984).

7.02.5.1.4 Advection–diffusion equation and reaction rates

Among the entropy sources, only terms involving similar tensorial ranks can be coupled in an isotropic medium, according to the Curie principle. The positivity of the entropy production [173] imposes three conditions, coupling tensors, vectors, and scalars:

$$\underline{\boldsymbol{\tau}} : \nabla \mathbf{v}_b \geq 0, \quad -\mathbf{q} \cdot \frac{\nabla T}{T} - \sum_i \mathbf{J}_i \cdot \nabla_T \mu_i \geq 0 \quad -\sum_i \Gamma_i \mu_i \geq 0 \qquad [174]$$

The first term relates tensors and we have already discussed its implications for the rheology in **Section 7.02.3.2**.

The second term relates vectors and we assume, in agreement with the general principle of nonequilibrium thermodynamics (de Groot and Mazur, 1984), that a phenomenological matrix M relates the thermodynamic fluxes $\mathbf{J} = \mathbf{J}_1 \ldots \mathbf{J}_i \ldots \mathbf{q}$ to the thermodynamic forces $\mathbf{X} = -\nabla_T \mu_1 \ldots -\nabla_T \mu_i \ldots -\nabla T/T$:

$$\begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \ldots \\ \mathbf{q} \end{pmatrix} = -\begin{pmatrix} m_{11} & m_{12} & \ldots & m_{1q} \\ m_{21} & m_{22} & \ldots & m_{2q} \\ \ldots & \ldots & \ldots & \ldots \\ m_{q1} & m_{q2} & \ldots & m_{qq} \end{pmatrix} \begin{pmatrix} \nabla_T \mu_1 \\ \nabla_T \mu_2 \\ \ldots \\ \nabla T/T \end{pmatrix} \qquad [175]$$

This linear relationship implies that the term of vectorial rank (with superscript υ), in the entropy source, $TH_S^{(\upsilon)}$, appears as

$$TH_S^{(\upsilon)} = \mathbf{X}^t M \mathbf{X} = \mathbf{X}^t \frac{M + M^t}{2} \mathbf{X} \qquad [176]$$

According to the second law of thermodynamics, the symmetrical part of the matrix M, $(M + M^t)/2$, must be positive definite, that is, the right-hand side of equation [176] must be positive for any vectors $\mathbf{X}$.

At microscopic scale, a process and its reverse occur at the same rate. A consequence, known as the Onsager reciprocal relations, is the existence of symmetry or antisymmetry between m_{ij} and m_{ji} (Onsager, 1903–76). A general discussion can be found in, for example, de Groot and Mazur (1984) or Woods (1975). When the forces are independent of the velocities, the matrix M must be symmetrical, $m_{ij} = m_{ji}$.

In the general case, the transport of heat by concentration gradients (the Dufour effect (Dufour, 1832–92)) or the transport of concentration by temperature gradients (the Soret effect (Soret, 1827–90)) is possible. In many situations, these cross effects are small and we will assume that the matrix M does not couple thermal and compositional effects (the last row and column of M are zero except for $m_{qq}/T = k$, the thermal conductivity).

Even without coupling between thermal and compositional effects, chemical diffusion in a multicomponent system remains difficult to discuss in the most general case (the positive definiteness of a symmetrical matrix is not a very strong constraint). We therefore restrict our study to a simple two-component system where

$$\begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \end{pmatrix} = -\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} \nabla_T \mu_1 \\ \nabla_T \mu_2 \end{pmatrix} \qquad [177]$$

For such a simple case, the sum of the fluxes must cancel (see eqn [157]), and since the Onsager relations impose the symmetry of the matrix, the coefficients m_{ij} must verify

$$m_{11} + m_{21} = m_{12} + m_{22} = m_{12} - m_{21} = 0 \qquad [178]$$

Only one coefficient, for example, $m = m_{11}$, can be freely chosen, and the fluxes can be written as

$$\begin{aligned} \mathbf{J}_1 &= m \nabla_T (\mu_2 - \mu_1) \\ \mathbf{J}_2 &= m \nabla_T (\mu_1 - \mu_2) \end{aligned} \qquad [179]$$

and the second law requires $m > 0$. If the component 1 is in small quantity (the solute) and the component 2 is in large quantity (the solvent, with $\nabla_T(\mu_2) = 0$), we can easily track the evolution of solute concentration C_1. Its chemical diffusion flux is $\mathbf{J}_1 = -m \nabla_T \mu_1$, and according to eqn [158], its concentration

$$\overline{\rho} \left(\frac{\partial C}{\partial t} + \mathbf{v}_b \cdot \nabla C\right) = \nabla \cdot (m \nabla_T \mu) \qquad [180]$$

where the subscript 1 has been omitted and the source term Γ_1 neglected.

For a solute, the chemical potential is a standard chemical potential μ_0 plus a mixing term expressing the entropy gain (configurational entropy associated with the increased disorder) made by dispersing the solute into the solvent, of the form $RT \log a(C)$ (for crystalline solids, the activity $a(C)$ of the mixing term can be complex since it depends on the number and multiplicity of crystallographic sites (Spear, 1993), but we

just need to know that it is related to C). In a domain where the average density and the standard potential remain uniform, the advection–diffusion equation is obtained:

$$\frac{\partial C}{\partial t} + \mathbf{v}_b \cdot \nabla C = \nabla \cdot (D \nabla C) \qquad [181]$$

with a diffusion coefficient $D = m/\overline{\rho}(\partial\mu/\partial C)$ most likely T-dependent. The negative linear relationship between chemical diffusion flux $\mathbf{J}$ (now in mol m^{-2} s^{-1}) and concentration gradient $\mathbf{J} = -D\nabla C$ (with D in m^2 s^{-1}) is called the first Fick's law (Fick, 1829–1901).

When a component is present in two domains separated by a compositional interface, its standard chemical potential μ_0 is generally discontinuous. In this case, the gradient of the chemical potential at constant T, $\nabla_T \mu$ is a mathematical distribution that contains a term $\nabla_T \mu_0$, infinite on the compositional interface. This discontinuity drives an infinitely fast diffusion of the solute component across the interface until the equilibrium and the continuity of the total chemical potential $\mu_0 + RT \log a(C)$ are restored. The concentration ratio of C (or partition coefficient of C) must therefore verify

$$\frac{a(C)^+}{a(C)^-} = \exp\left(-\frac{\mu_0^+ - \mu_0^-}{RT}\right) \qquad [182]$$

where $\mu_0^+ - \mu_0^- = [\mu_0]$ is the jump of standard potential across the interface. This equation corresponds to the general rule of chemical equilibrium.

The last entropy source in eqn [174], $\sum_i \Gamma_i \mu_i$, relates two scalars (production rates and chemical potentials). In a mixture of i components involving k stable atomic species, the conservation of these atomic species implies that only $r = i - k$ linearly independent reactions exist. Let n_i^j be the stoichiometric coefficient of the component i in the jth $= (1 \ldots r)$ chemical reaction with reaction rate, Γ_j. We can express Γ_i as

$$\Gamma_i = \sum_{j=1\ldots r} n_i^j \Gamma_j \qquad [183]$$

and the second law imposes

$$-\sum_{j=1\ldots r} \Gamma_j \sum_i n_i^j \mu_i \geq 0 \qquad [184]$$

The positivity of the entropy source is satisfied if the kinetic rates of the jth $= 1 \ldots r$ chemical reaction are proportional to their chemical affinities, $-\Delta G_j = -\sum_i n_i^j \mu_i$, with positive reaction rate factors $\mathcal{R}_j$:

$$\Gamma_j = -\mathcal{R}_j \Delta G_j \qquad [185]$$

In the case of exact thermodynamic equilibrium, $\Delta G_i = 0$, the reaction rate vanishes and the second law is of course satisfied.

Chemical reaction rates are however very rarely simply proportional to the affinities and the $\mathcal{R}_j$ are likely some complex, but positive, functions of P, T and concentrations C_i. Indeed, in the context of chemical reactions, the kinetics of reactions are often approximated by phenomenological laws based on the law of mass action rather than on affinities (e.g., the reaction rates are rather related to $\exp(-\Delta G_j/RT) - 1$ than to $-\Delta G_j$ (see Lasaga, 1998; Rudge et al., 2011)).

In the same way as we defined the affinity $-\Delta G_j$ of the reaction j, we can define its enthalpy $\Delta\mathcal{H}_j = \sum_i n_i^j \mathcal{H}_i$ (see de Groot and Mazur, 1984). The enthalpy exchange term of the energy equation [165] $\sum_i \Gamma_i \mathcal{H}_i$ can also be written as $\sum_j \Gamma_j \Delta\mathcal{H}_j$, which represents the products of the reaction rates and the reaction enthalpies. Various phase changes take place in the mantle, most notably at 410 and 660 km depth. Their effects on mantle convection have been studied by various authors and will be discussed in Section 7.02.6.6.

7.02.5.1.5 Conservation properties of the advection–diffusion equation

We now make the hypothesis that the evolution of concentration of a solute in the convective fluid is controlled by the advection–diffusion equation [181] and that this solute is not involved in any chemical reaction, $\Gamma = 0$. For simplicity, we assume that the barycentric flow is incompressible ($\nabla \cdot \mathbf{v}_b = 0$, C can therefore be a concentration per unit volume or per unit mass) and the diffusion coefficient D is a constant. The fluid and the solute cannot escape the domain Ω; the normal velocity and normal diffusive flux are thus zero on the boundaries of the domain, that is, $\mathbf{v}_b \cdot \mathbf{n} = 0$ and $D\nabla C \cdot \mathrm{n}_b = 0$ on the surface $\sum$ with normal vector $\mathbf{n}$.

First, it is obvious that when integrated over the total domain Ω and with the divergence theorem, the advection–diffusion equation [181] implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega C \mathrm{d}V = -\int_\Sigma (C\mathbf{v}_b - D\nabla C) \cdot \mathrm{d}\mathbf{S} = 0 \qquad [186]$$

The initial heterogeneity does not disappear, it is just redistributed through time.

To understand how the heterogeneity is redistributed, we can express the evolution of the concentration variance. The variance is related to $C^2 - \overline{C}^2$ where $\overline{C}$ is the average concentration. We can get an evolution equation for C^2 by multiplying eqn [181] by $2C$ and we get after some algebra

$$\frac{\mathrm{D}C^2}{\mathrm{D}t} = D\nabla^2 C^2 - 2D|\nabla C|^2 \qquad [187]$$

This expression when integrated over the closed volume Ω implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega C^2 \mathrm{d}V = -2D\int_\Omega |\nabla C|^2 \mathrm{d}V \qquad [188]$$

Since the right side is always negative, the variance must continuously decrease until $|\nabla \mathrm{C}| = 0$, which corresponds to a state of complete homogenization.

If the initial concentration is $C = 1$ in some domain and $C = 0$ elsewhere, a perfect homogenization is achieved after a time t if the concentration is everywhere $C = \overline{C}$, the average concentration. Since the average mixing rate is proportional to the diffusion D, eqn [188], we note, however, that a nondiffusive flow does not homogenize at all. A nondiffusive flow just stirs the heterogeneities. When there is no diffusion, the initial heterogeneity is stirred and stretched, but the local concentrations remain, for all time, either $C = 1$ or $C = 0$, but never an intermediate value (see Figure 5)

In the case of the Earth's mantle, the solid-state diffusion coefficients are all very low (see Table 2), and many studies have totally neglected chemical diffusion. We see, at face value, that these models are not really homogenizing, only stirring the heterogeneities. Without diffusion, a chemical

heterogeneity (e.g., a piece of subducted oceanic crust) will forever remain the same petrologic heterogeneity, and only its shape will change.

Since the mixing rate is related to the compositional gradient [188], we should discuss the evolution of this gradient. We multiply eqn [181] by the operator $2\nabla C\cdot\nabla$ to obtain

$$\frac{\mathrm{D}|\nabla C|^2}{\mathrm{D}t} = -2\nabla C\cdot\underline{\dot{\boldsymbol{\varepsilon}}}\cdot\nabla C + 2D\left[\nabla\cdot\left(\nabla^2 C\nabla C\right) - \left(\nabla^2 C\right)^2\right] \quad [189]$$

which can be integrated as

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega |\nabla C|^2\,\mathrm{d}V = -2\int_\Omega \nabla C\cdot\underline{\dot{\boldsymbol{\varepsilon}}}\cdot\nabla C\,\mathrm{d}V - 2D\int_\Omega \left(\nabla^2 C\right)^2\mathrm{d}V \quad [190]$$

The rate of gradient production is related to the flow properties through the strain rate tensor $\underline{\dot{\boldsymbol{\varepsilon}}}$ and to the diffusion. The diffusion term is negative and decreases the sharpness of compositional gradients.

The term related to the flow properties through the strain tensor (first term of the right side of eqn [190]) could in principle be either positive or negative. However, as time evolves, this term must become positive. The strain rate tensor has locally three principal axes and three principal strain rates, the sum of them being zero since the flow is incompressible. The stretched heterogeneities become elongated along the direction of the maximum principal strain rate and the concentration gradients reorient themselves along the minimum, and negative, principal strain rate. The term under the first integral on the right side of eqn [190] is thus of the order of $+|\dot{\boldsymbol{\varepsilon}}_{\min}||\nabla C|^2$ ($\dot{\boldsymbol{\varepsilon}}_{\min}$ is the local, negative eigenvalue of the strain rate tensor). Stirring is thus the source of production of concentration gradient (see **Figure 5**).

We can now understand the interplay between advection and diffusion. Even when the diffusion coefficient D is vanishingly small in eqn [190], the stirring of the flow by convection will enhance the concentration gradients until the average diffusion term, proportional to the square of concentration gradients, will become large enough (see eqn [188]), for a rapid decrease of the concentration variance. We illustrate this behavior in the next two subsections by choosing a simple expression for the strain rate and computing the evolution of concentration through time.

7.02.5.1.6 Laminar and turbulent stirring

The efficiency of mixing, mostly controlled by stirring, is therefore related to the ability of the flow to rapidly reduce the thickness of heterogeneities (Olson et al., 1984). In this section, we set aside diffusion and discuss the stirring properties of a flow (see also Tackley, this volume). Let us consider a vertical piece of heterogeneity of width $2d_0$ and height $2l_0$ ($l_0 \gg d_0$) in a simple shear flow $v_x = \dot{\boldsymbol{\varepsilon}}z$, with constant $\dot{\boldsymbol{\varepsilon}}$. Its top and bottom ends are at $(0, l_0)$ and $(0, -l_0)$ and they will be advected to $(\dot{\boldsymbol{\varepsilon}}tl_0, l_0)$, $(-\dot{\boldsymbol{\varepsilon}}tl_0, -l_0)$ after a time t. As the heterogeneity length increases as $2l_0\sqrt{1+\dot{\boldsymbol{\varepsilon}}^2t^2}$, mass conservation implies that its half width $d(t)$ decreases as

$$d(t) = \frac{d_0}{\sqrt{1+\dot{\boldsymbol{\varepsilon}}^2t^2}} \quad [191]$$

On the contrary, in a pure shear flow, $v_z = \mathrm{d}z/\mathrm{d}t = z\dot{\boldsymbol{\varepsilon}}$ with constant $\dot{\boldsymbol{\varepsilon}}$, the length of the heterogeneity would increase as $l = l_0\exp(\dot{\boldsymbol{\varepsilon}}t)$ and its width would shrink as

$$d(t) = d_0\exp(-\dot{\boldsymbol{\varepsilon}}t) \quad [192]$$

Table 2 Homogenization times for helium and uranium assuming a heterogeneity of initial thickness $2d_0 = 7$ km and a strain rate of 5×10^{-16} s^{-1}

	Uranium	*Helium*
D (m^2 s^{-1})	10^{-19}	10^{-13}
t_0 (Ma)	3.88×10^{13}	3.88×10^{7}
t_L (Ma)	3.60×10^{5}	3.60×10^{3}
t_T (Ma)	1575	699

The diffusion coefficients for U and He are from Hofmann and Hart (1978) and Trull and Kurtz (1993).

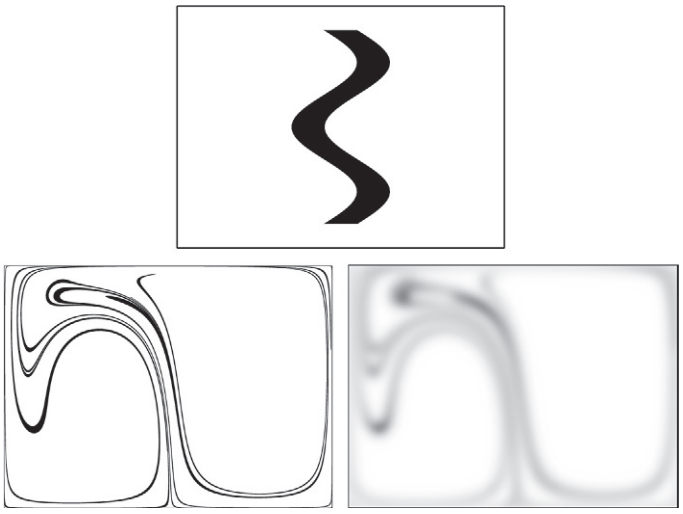


Figure 5 An initial heterogeneity (top) is introduced at $t = 0$ into a time-dependent convection cell. Without diffusion, $D = 0$ (bottom left), the heterogeneity is stirred by convection and then stretched on the form of thin ribbons. However, the variance of the heterogeneity concentration remains constant. It is only with diffusion, $D \neq 0$ (bottom right), that a real homogenization occurs with a decrease of the heterogeneity variance.

Such a flow is said to induce turbulent stirring. This is unfortunate terminology because turbulent stirring can occur in a creeping flow with $Re=0$. Mantle convection is not turbulent but it generates turbulent stirring.

Chaotic mixing flows have globally turbulent stirring properties and the qualitative idea that highly time-dependent convection with high Rayleigh number mixes more efficiently than low Rayleigh number convection is often true (Coltice and Schmalzl, 2006; Schmalzl et al., 1996). However, steady 3-D flows can also induce turbulent mixing. This surprising phenomenon called Lagrangian chaos is well illustrated for some theoretical flows (Dombre et al., 1986) and for various simple flows (Ottino, 1989; Toussaint et al., 2000). For example, in a steady flow under an oceanic ridge offset by a transform fault, the mixing is turbulent (Ferrachat and Ricard, 1998).

7.02.5.1.7 Diffusion in Lagrangian coordinates

In Section 7.02.5.1.5, we discussed the mixing properties from a Eulerian viewpoint. We can also understand the interplay between diffusion and stretching (stirring) by adopting a Lagrangian viewpoint (Kellogg and Turcotte, 1987; Ricard and Coltice, 2004), that is, by solving the advection–diffusion equations in a coordinate frame that follows the deformation.

Let us consider a strip of thickness $2d_0$ with an initial concentration C_0 embedded in an infinite matrix of concentration C_∞. In the absence of motion, the solution of the advection–diffusion equation [181] can be expressed using the error function, and the time-dependent concentration $C(x, t)$ is given by

$$\frac{C(x,t)-C_\infty}{C_0-C_\infty}=\frac{1}{2}\left[\mathrm{erf}\left(\frac{d_0-x}{2\sqrt{Dt}}\right)+\mathrm{erf}\left(\frac{d_0+x}{2\sqrt{Dt}}\right)\right] \quad [193]$$

where x is a coordinate perpendicular to the strip and is zero at its center.

The concentration at the center of the strip ($x=0$) is

$$\frac{C(0,t)-C_\infty}{C_0-C_\infty}=\mathrm{erf}\left(\frac{d_0}{2\sqrt{Dt}}\right) \quad [194]$$

and the concentration decreases by a factor of about 2 in the diffusive time

$$t_0\approx\frac{d_0^2}{D} \quad [195]$$

($\mathrm{erf}(1/2)$ is not far from 1/2). The time needed to homogenize a 7 km thick piece of oceanic crust introduced into a motionless mantle is extremely long (see Table 2); even the relatively mobile helium would be frozen in place since the Earth formed as it would only have migrated around 50 cm.

However, this idea of a $\sqrt{t}$ diffusion is faulty since the flow stirs the heterogeneity and increases compositional gradients (see eqn [190]), which in turn accelerates the mixing process (e.g., eqn [188]). Assuming that the problem remains two-dimensional enough so that diffusion only occurs perpendicular to the deforming heterogeneity, let $d(t)$ with time derivative $\dot{d}(t)$ be half the thickness of the strip containing a chemical heterogeneity. The velocity perpendicular to the strip would locally be at first order

$$v_x=x\frac{\dot{d}(t)}{d(t)} \quad [196]$$

(each side of the stripe, at $x=\pm d(t)$ moves at $\pm\,\dot{d}(t)$).

We can choose as a new space variable $\tilde{x}=xd_0/d(t)$, in such a way that the Lagrangian coordinate $\tilde{x}$ will vary between the fixed values $-d_0$ and d_0. The diffusion equation becomes

$$\left(\frac{\partial C}{\partial t}\right)_{\tilde{x}}=D\left(\frac{d_0}{d(t)}\right)^2\frac{\partial^2 C}{\partial\tilde{x}^2} \quad [197]$$

where the partial time derivative is now computed at constant $\tilde{x}$. We see that the advection–diffusion equation has been turned into a pure diffusive equation where the diffusivity D has been replaced by $D(d_0/d(t))^2$. This equivalent diffusivity is larger than D and increases with time as $d(t)$ decreases.

To solve analytically equation [197], it is appropriate to rescale the time variable by defining $\tilde{t}=F(t)$ with

$$F(t)=\int_0^t\left(\frac{d_0}{d(u)}\right)^2\mathrm{d}u \quad [198]$$

and the resulting advection–diffusion equation in Lagrangian coordinates becomes the simple diffusion equation with constant diffusivity where t and x are replaced by $\tilde{t}$ and $\tilde{x}$. Its solution can be deduced from eqn [193]. For example, the concentration at the center of the deformable strip varies like

$$\frac{C(0,t)-C_\infty}{C_0-C_\infty}=\mathrm{erf}\left(\frac{d_0}{2\sqrt{DF(t)}}\right) \quad [199]$$

and the concentration diminishes in amplitude by a factor of 2 after a time t that satisfies

$$F(t)\approx\frac{d_0^2}{D} \quad [200]$$

To perform a numerical application, let us consider that the flow is either a simple shear, eqn [191], or a pure shear deformation, eqn [192]. Computing $F(t)$ from equation [198] is straightforward and, assuming $\dot{\boldsymbol{\epsilon}}t\gg 1$, we get from eqn [200] the homogenization times

$$t_\mathrm{L}\approx\frac{3^{1/3}d_0^{2/3}}{\dot{\boldsymbol{\epsilon}}^{2/3}D^{1/3}} \quad [201]$$

and

$$t_\mathrm{T}\approx\frac{1}{\dot{\boldsymbol{\epsilon}}}\log\frac{d_0^2\dot{\boldsymbol{\epsilon}}}{D} \quad [202]$$

for the laminar and turbulent case, respectively. For the same oceanic crust of initial thickness 7 km, we get homogenization times of about 0.70 byr for He and 1.57 byr for U if we use the pure shear mechanism and assume rather arbitrarily that $\dot{\boldsymbol{\epsilon}}=5\times10^{-16}\ \mathrm{s}^{-1}$ (this corresponds to a typical plate velocity of 7 cm year^{-1} over a plate length of 5000 km). Although He and U have diffusion coefficients 6 orders of magnitude apart, their residence times in a piece of subducted oceanic crust only differ by a factor ≈ 2 (see Table 2).

The use of tracers to simulate the evolution of chemical properties in the mantle is our best method since solid-state diffusion is too slow to be efficiently accounted for in a numerical simulations (e.g., Tackley and Xie, 2002; van Keken et al., 2002). However, by using tracers, we do not necessarily take

into account that some of them may represent points that have been stretched so much that their initial concentration anomalies have completely diffused into the background. In other words, even if diffusion seems negligible, diffusion will erase all heterogeneities after a finite time that is mostly controlled by the stirring properties of the flow.

7.02.5.2 Fluid Dynamics of Two-Phase Flows

Up to now, in all of Section 7.02.5.1, all components were mixed in a single phase. However, another important geophysical application occurs when the multicomponents belong to different phases. This case can be illustrated with the dynamics of partial melt in a deformable compacting matrix. Partial melts are obviously present under ridges and hot spots, but they may also be present in the middle and deep mantle (Bercovici and Karato, 2003; Williams and Garnero, 1996) and they were certainly more frequent when the Earth was younger (Labrosse et al., 2007). We discuss the situation where two phases, fluid and matrix, can interact. In contrast to Section 7.02.5.1, where the proportion and velocity of each component in solution were defined everywhere at a microscopic level, in a partial melt aggregate, the local velocity at a microscopic level is either the velocity of a matrix grain, $\tilde{\mathbf{v}}_m$, or the interstitial velocity of the melt, $\tilde{\mathbf{v}}_f$.

We assume that the two phases are individually homogeneous and incompressible (various extensions of this theory are possible, for compressible phases (e.g., Bercovici and Michaut, 2010; Michaut et al., 2013)), when one phase is void (e.g., Bercovici and Ricard, 2003; Ricard and Bercovici, 2003) and when one phase is only manifested by its surface energy (e.g., Hier-Majumder et al., 2006). They have densities ρ_f and ρ_m and Newtonian rheologies with viscosities η_f and η_m. They are isotropically mixed and connected. Their volume fractions are ϕ (the porosity) and $1-\phi$. The rate of magma melting or freezing is $\Delta\Gamma$ (in kg m^{-3} s^{-1}). Although the two phases have very different physical properties, we will require the equations to be material-invariant until we need to use numerical values. This means that swapping f and m, ϕ and $1-\phi$, and $\Delta\Gamma$ and $-\Delta\Gamma$ must leave the equations unchanged. This rule is both a physical requirement and a strong guidance in establishing the general equations (Bercovici et al., 2001a).

We make the hypothesis that there is a mesoscopic size of volume δV, which includes enough grains and interstitial fluid that averaged and continuous quantities can be defined. Classical fluid dynamics also has its implicit averaging volume δV that must contain enough atoms that quantum effects are negligible, but what is needed here is a much larger volume. This averaging approach remains meaningful because the geophysical macroscopic phenomenon that we want to understand (say, melting under ridges) has characteristic sizes large compared to those of the averaging volume (say, a few cm^3) (Bear, 1988).

To do the averaging, we define at microscopic level a function θ that takes the value 1 in the interstitial fluid and the value 0 in the matrix grain. Mathematically, this function is rather a distribution and it has a very convolved topology. From it, we can define first the porosity (volume fraction of fluid) ϕ and then the fluid and matrix averaged velocities, $\mathbf{v}_f$ and $\mathbf{v}_m$ (Bercovici et al., 2001a), by

$$\phi = \frac{1}{\delta V}\int_{\delta V} \theta \, dV \qquad [203]$$

$$\phi \mathbf{v}_f = \frac{1}{\delta V}\int_{\delta V} \theta \tilde{\mathbf{v}}_f \, dV, \quad (1-\phi)\mathbf{v}_m = \frac{1}{\delta V}\int_{\delta V} (1-\theta)\tilde{\mathbf{v}}_m \, dV \qquad [204]$$

7.02.5.2.1 Mass conservation for matrix and fluid

Having defined the average quantities, the derivation of the two mass conservation equations is fairly standard (Bercovici et al., 2001a; McKenzie, 1984). They are

$$\frac{\partial \phi}{\partial t} + \nabla\cdot[\phi \mathbf{v}_f] = \frac{\Delta\Gamma}{\rho_f} \qquad [205]$$

$$-\frac{\partial \phi}{\partial t} + \nabla\cdot[(1-\phi)\mathbf{v}_m] = -\frac{\Delta\Gamma}{\rho_m} \qquad [206]$$

We get the same equations as in eqn [150] except that we refer to ϕ, $1-\phi$, and $\Delta\Gamma$ instead of ϕ_1, ϕ_2, and Γ_1. When averaged, the mass conservation equations of two separated phase take the same form as the mass conservation equations of the two components in a solution.

We define an average and a difference quantity for any general variable, q, by

$$\overline{q} = \phi q_f + (1-\phi) q_m, \quad \Delta q = q_m - q_f \qquad [207]$$

The velocity $\overline{\mathbf{v}}$ is volume-averaged and is different from the barycentric velocity (eqn [152]),

$$\mathbf{v}_b = (\phi\rho_f \mathbf{v}_f + (1-\phi)\rho_m \mathbf{v}_m)/\overline{\rho} = \overline{v} + \phi(1-\phi)\Delta v \Delta\rho/\overline{\rho}$$

By combining the fluid and matrix mass conservation equations, we get the total mass conservation equation

$$\frac{\partial \overline{\rho}}{\partial t} + \nabla\cdot(\overline{\rho}\mathbf{v}_b) = 0 \qquad [208]$$

(as before, eqn [153]) and the time rate of change in volume during melting

$$\nabla\cdot\overline{\mathbf{v}} = \Delta\Gamma\frac{\Delta\rho}{\rho_f\rho_m} \qquad [209]$$

7.02.5.2.2 Momentum conservation of matrix and fluid

Total momentum conservation, that is, the balance of the forces applied to the mixture, is

$$\nabla\cdot\underline{\overline{\boldsymbol{\tau}}} - \nabla\overline{P} + \overline{\rho}\mathbf{g} = 0 \qquad [210]$$

We have considered that the only force is due to gravity, although surface tension between the two phases could also be introduced (Bercovici and Ricard, 2003; Bercovici et al., 2001a). In this equation, $\overline{P}$, $\underline{\overline{\boldsymbol{\tau}}}$, and $\overline{\rho}$ are the average pressure, stress, and density, respectively. The equation is not surprising and looks identical to its counterpart for a multicomponent solution [160]. However, the average pressure and stresses $\overline{P} = \phi P_f + (1-\phi)P_m$ and $\underline{\overline{\boldsymbol{\tau}}} = \phi\underline{\boldsymbol{\tau}}_f + (1-\phi)\underline{\boldsymbol{\tau}}_m$ are now the sum of two separate contributions, from two separate phases having most likely very different rheologies and different pressures. Hypothesizing that the two phases are subject to the same pressure does not rest on any physical justification. It certainly cannot hold if surface tension is present as the Young–Laplace condition implies in the static case, a pressure difference between phases related to the curvature of their interface. We will show later that the pressure difference between phases ΔP controls the rate of porosity change.

We split the total momentum equation into two equations one for the fluid and one for the matrix:

$$\nabla[\phi P_f] + \phi\rho_f \mathbf{g} + \nabla.\left[\phi\underline{\overline{\tau}}_f\right] + \mathbf{h}_f = 0 \quad [211]$$

$$-\nabla[(1-\phi)P_m] + (1-\phi)\rho_m \mathbf{g} + \nabla.\left[(1-\phi)\underline{\overline{\tau}}_m\right] + \mathbf{h}_m = 0 \quad [212]$$

where $\mathbf{h}_f$ and $\mathbf{h}_m$ satisfy $\mathbf{h}_f + \mathbf{h}_m = 0$ and represent the interaction forces acting on the fluid and on the matrix, across the interfaces separating the two phases. Because of the complexities of the interfaces, these two interaction forces must be parameterized in some way.

The simplest contribution to the interfacial forces that preserves Galilean invariance is a Darcy-like term $c\Delta\mathbf{v} = c(\mathbf{v}_m - \mathbf{v}_f)$ (Drew and Segel, 1971; McKenzie, 1984) (Darcy, 1803–58). The interaction coefficient c is related to permeability that is itself a function of porosity (Bear, 1988). A symmetrical form compatible with the usual Darcy term is (see Bercovici et al., 2001a)

$$c = \frac{\eta_f \eta_m}{k_0\left[\eta_f(1-\phi)^{n-2} + \eta_m \phi^{n-2}\right]} \quad [213]$$

where the permeability of the form $k_0\phi^n$ was used (this is often called the Kozeny–Carman law with $n \approx 2$–3 (Carman, 1939)). Assuming $n = 2$ and $\eta_m \gg \eta_f$, the interaction coefficient becomes a constant, $c = \eta_f/k_0$.

In the absence of gravity and when the pressures are uniform and equal, no motion should occur even in the presence of nonuniform porosity. In this situation where $\Delta\mathbf{v} = \underline{\overline{\tau}}_f = \underline{\overline{\tau}}_m = 0$ and where P is uniform, $P = P_f = P_m$, the force balances are $-P\nabla\phi + \mathbf{h}_f = -P\nabla(1-\phi) + \mathbf{h}_m = 0$. Therefore, the interface forces $\mathbf{h}_f$ and $\mathbf{h}_m$ must also include $P\nabla\phi$ and $P\nabla(1-\phi)$ when the two pressures are equal. This led Bercovici and Ricard (2003) to write the interaction terms as

$$\begin{aligned} \mathbf{h}_f &= c\Delta\mathbf{v} + P_f\nabla\phi + \omega\Delta P\nabla\phi \\ \mathbf{h}_m &= -c\Delta\mathbf{v} + P_m\nabla(1-\phi) + (1-\omega)\Delta P\nabla\phi \end{aligned} \quad [214]$$

with $0 \le \omega \le 1$. These expressions verify $\mathbf{h}_f + \mathbf{h}_m = 0$, are Galilean- and material-invariant, and allow equilibrium of a mixture with nonuniform porosity but uniform and equal pressures (see also McKenzie, 1984).

At microscopic level, the matrix–melt interfaces are not sharp discontinuities but correspond to layers (called 'selvage' layers) of disorganized atom distributions. The coefficient $0 < \omega < 1$ controls the partitioning of the pressure jump (and potentially of the surface tension) between the two phases (Bercovici and Ricard, 2003) and represents the fraction of the volume-averaged surface force exerted on the fluid phase. The exact value of ω is related to the microscopic behavior of the two phases (molecular bond strengths and thickness of the interfacial selvage layers) and measures the extent to which the microscopic interface layer is embedded in one phase more than the other. The only general physical constraints that we have are that ω must be zero when the fluid phase disappears (when $\phi = 0$) and when the fluid phase becomes unable to sustain stresses (when $\eta_f = 0$). A symmetrical form like

$$\omega = \frac{\phi\eta_f}{\phi\eta_f + (1-\phi)\eta_m} \quad [215]$$

satisfies these conditions. For mantle conditions, $\eta_f \ll \eta_m$ and hence, $\omega \approx 0$.

To summarize, general expressions for the equations of fluid and matrix momentum conservation are (Bercovici and Ricard, 2003)

$$-\phi[\nabla P_f - \rho_f \mathbf{g}] + \nabla.\left[\phi\underline{\boldsymbol{\tau}}_f\right] + c\Delta\mathbf{v} + \omega\Delta P\nabla\phi = 0 \quad [216]$$

$$-(1-\phi)[\nabla P_m - \rho_m \mathbf{g}] + \nabla.\left[(1-\phi)\underline{\boldsymbol{\tau}}_m\right] - c\Delta\mathbf{v} + (1-\omega)\Delta P\nabla\phi = 0 \quad [217]$$

The relationship between stress and velocities does not include an explicit bulk viscosity term (Bercovici et al., 2001a), and for each phase j, the deviatoric stress is simply

$$\underline{\boldsymbol{\tau}}_j = \eta_j\left(\nabla\mathbf{v}_j + \left[\nabla\mathbf{v}_j\right]^t - \frac{2}{3}\nabla\cdot\mathbf{v}_j\underline{\mathbf{I}}\right) \quad [218]$$

where j stands for f or m. There is no difference in constitutive relations for the isolated component and the component in the mixture.

7.02.5.2.3 Energy conservation for two-phase flows

In the case where surface energy and entropy exist on interfaces, the conservation of energy deserves more care (Bercovici and Ricard, 2003; Bercovici et al., 2001a; Hier-Majumder et al., 2006; Sramek et al., 2007). Otherwise, the global conservation is straightforward and can be expressed by the following equation where the left side represents the temporal change of internal energy content in a fixed control volume and the right side represents the different contributions to this change, namely, internal heat sources, loss of energy due to diffusion, advection of energy, and rate of work of both surface and body forces,

$$\begin{aligned} \frac{\partial}{\partial t}[\phi\rho_f\mathcal{U}_f + (1-\phi)\rho_m\mathcal{U}_m] &= \overline{\rho}\overline{H} - \nabla\cdot\mathbf{q} - \nabla\cdot[\phi\rho_f\mathcal{U}_f\mathbf{v}_f \\ &+ (1-\phi)\rho_m\mathcal{U}_m\mathbf{v}_m] + \nabla\cdot\left[-\phi P_f\mathbf{v}_f - (1-\phi)P_m\mathbf{v}_m + \phi\mathbf{v}_f\cdot\underline{\boldsymbol{\tau}}_f\right. \\ &\left. + (1-\phi)\mathbf{v}_m\cdot\underline{\boldsymbol{\tau}}_m\right] + \phi\mathbf{v}_f\cdot\rho_f\mathbf{g} + (1-\phi)\mathbf{v}_m\cdot\rho_m\mathbf{g} \end{aligned} \quad [219]$$

The last equation is manipulated in the standard way using the mass and momentum equations. Because the two phases are incompressible, their internal energies are simply $d\mathcal{U}_f = C_f dT$ and $d\mathcal{U}_m = C_m dT$. After some algebra, we get

$$\begin{aligned} \phi\rho_f C_f\frac{D_f T}{Dt} + (1-\phi)\rho_m C_m\frac{D_m T}{Dt} &= -\nabla\cdot\mathbf{q} - \Delta P\frac{D_\omega\phi}{Dt} \\ &+ \Delta\mathcal{H}\Delta\Gamma + \Psi + \overline{\rho}\overline{H} \end{aligned} \quad [220]$$

where Ψ is the rate of deformational work

$$\Psi = \phi\nabla\mathbf{v}_f : \underline{\boldsymbol{\tau}}_f + (1-\phi)\nabla\mathbf{v}_m : \boldsymbol{\tau}_m + c(\Delta v)^2 \quad [221]$$

It contains the dissipation terms of each phase plus a term related to the friction between the two phases. The fundamental derivatives are defined by

$$\frac{D_j}{Dt} = \frac{\partial}{\partial t} + \mathbf{v}_j\cdot\nabla \quad [222]$$

where $\mathbf{v}_j$ is to be substituted with the appropriate velocity $\mathbf{v}_f$, $\mathbf{v}_m$, or $\mathbf{v}_\omega$ with

$$\mathbf{v}_\omega = \omega\mathbf{v}_f + (1-\omega)\mathbf{v}_m \quad [223]$$

In contrast to Section 7.02.5.1, it would not make much sense to try to keep the equations in terms of an average velocity like $\mathbf{v}_b$ plus some minor diffusion terms. Here, the two components may have very different velocities (e.g., a weak downward compaction of the matrix with a fast upward migration of magma), and we have to define various material derivatives.

Since ω represents a partitioning of pressure jump, it is not surprising to find the velocity $\mathbf{v}_\omega$ (included in D_ω/Dt) in the work term related to this pressure jump. Associated with this partitioning factor, we can also introduce interface values, q^ω, that we will use later. Any quantity $\Delta q = q_m - q_f$ can also be written as $(q_m - q^\omega) - (q_f - q^\omega)$. When the property jump is embedded entirely in the matrix ($\omega = 0$), there should be no jump within the fluid and we must have $q^\omega = q_f$. Reciprocally, when $\omega = 1$, we should have $q^\omega = q_m$. This prompts us to define interface values by

$$q^\omega = (1-\omega)q_f + \omega q_m \qquad [224]$$

Notice in the expressions of the interface velocity $\mathbf{v}_\omega$, eqn [223], and interface value q^ω, eqn [224], that ω and $1-\omega$ are interchanged.

The right side of eqn [220] contains two new expressions in addition to the usual terms (heat production, diffusion, and deformational work). The first term includes the changes in porosity $D_\omega\phi/Dt$ times the difference in pressures between phases, ΔP. The other term contains the difference in the specific enthalpies $\Delta\mathcal{H} = \mathcal{H}_m - \mathcal{H}_f$ where the enthalpy of phase j is defined by $\mathcal{H}_j = \mathcal{U}_j + P_j/\rho_j$. A similar term was found for components reacting in a solution [162].

7.02.5.2.4 Entropy production and phenomenological laws

Entropy conservation is needed to constrain the pressure jump between phases and the melting rate. Starting from entropy conservation and following Sramek et al. (2007)

$$\frac{\partial}{\partial t}[\phi\rho\mathcal{S}_f + (1-\phi)\rho_m\mathcal{S}_m] = -\nabla\cdot\mathbf{J}_\mathcal{S} + H_\mathcal{S} \qquad [225]$$

where $\mathbf{J}_\mathcal{S}$ is the total entropy flux and $H_\mathcal{S}$ is the internal entropy production, we compare the energy and the entropy equations [220] and [225] taking into account that, for each incompressible phase, $d\mathcal{S}_j = C_j\, dT/T = d\mathcal{U}_j/T$. After some algebra, one gets

$$\mathbf{J}_\mathcal{S} = \phi\rho_f\mathcal{S}_f\mathbf{v}_f + (1-\phi)\rho_m\mathcal{S}_m\mathbf{v}_m + \frac{\mathbf{q}}{T} \qquad [226]$$

$$TH_\mathcal{S} = -\frac{1}{T}\mathbf{q}\cdot\nabla T - \Delta P\frac{D_\omega\phi}{Dt} + \Delta\mu\Delta\Gamma + \Psi + \overline{\rho}\overline{H} \qquad [227]$$

where we have introduced the difference in chemical potentials between the two phases

$$\Delta\mu = \Delta\mathcal{H} - T\Delta S \qquad [228]$$

and $\Delta S = \mathcal{S}_m - \mathcal{S}_f$ is the change in specific entropies.

Following the standard procedure of nonequilibrium thermodynamics, we choose $\mathbf{q} = -k\nabla T$ and we assume that there is a linear relationship between the two thermodynamic fluxes, $D_\omega\phi/Dt$ and $\Delta\Gamma$, and the two thermodynamic forces $-\Delta P$ and $\Delta\mu$ since they have the same tensorial rank. We write

$$\begin{pmatrix} D_\omega\phi/Dt \\ \Delta\Gamma \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}\begin{pmatrix} -\Delta P \\ \Delta\mu \end{pmatrix} \qquad [229]$$

The matrix of phenomenological coefficients m_{ij} is positive definite and symmetrical by Onsager's theorem, $m_{12} = m_{21}$ (see Sramek et al., 2007). For a 2×2 matrix, it is rather simple to show that the positivity implies $m_{22} > 0$, $m_{11} > 0$, and $m_{11}m_{22} - m_{12}^2 > 0$ (positivity of the determinant).

The form of the phenomenological coefficients m_{ij} can be constrained through thought experiments. First, using mass conservations [205] and [206] and the definitions of $\mathbf{v}_\omega$ and ρ^ω, [223] and [224], we can combine equations [229] to get

$$\Delta P = -\frac{m_{22}}{m_{11}m_{22} - m_{12}^2}\left[(1-\omega)(1-\phi)\nabla\cdot\mathbf{v}_m - \omega\phi\nabla\cdot\mathbf{v}_f + \left(\frac{\rho^\omega}{\rho_f\rho_m} - \frac{m_{12}}{m_{22}}\right)\Delta\Gamma\right] \qquad [230]$$

In the limiting case where the two phases have the same density $\rho_f = \rho_m = \rho^\omega$, melting can occur with no motion, $\mathbf{v}_m = \mathbf{v}_f = 0$, and eqn [230] should therefore predict the equality of pressure between phases, $\Delta P = 0$. In this case, we must choose $m_{12}/m_{22} = 1/\rho_f = 1/\rho_m$. Let us consider now a situation of homogeneous isotropic melting where the melt has such a low viscosity that it cannot sustain viscous stresses and cannot interact with the solid by Darcy terms. For such an inviscid melt, $\omega = 0$, $\rho^\omega = \rho_f$ and $\mathbf{v}_\omega = \mathbf{v}_m$. In this case, since the melt can escape instantaneously, the matrix should not dilate, $\nabla\mathbf{v}_m = 0$, and thus, the two pressures should also be the same, $\Delta P = 0$. In this situation, all the terms in eqn [230] are 0 except for the term proportional to $\Delta\Gamma$. Thus, in the general case,

$$\frac{m_{12}}{m_{22}} = \frac{\rho^\omega}{\rho_f\rho_m} \qquad [231]$$

Using this condition and introducing two positive coefficients, $\zeta = m_{22}/(m_{11}m_{22} - m_{12}^2)$, and $\mathcal{R} = m_{22}$, we can recast eqn [229] as

$$\Delta P = -\zeta\left[\frac{D_\omega\phi}{Dt} - \frac{\rho^\omega}{\rho_f\rho_m}\Delta\Gamma\right] \qquad [232]$$

$$\Delta\Gamma = \mathcal{R}\left[\Delta\mu - \frac{\rho^\omega}{\rho_f\rho_m}\Delta P\right] \qquad [233]$$

The first equation establishes a general relation controlling the pressure drop between phases. The coefficient ζ that links the pressure jump between the two phases to the porosity changes in excess of the melting rate is in fact equivalent to a bulk viscosity as introduced in Section 7.02.3.2 (see also the summary in Section 7.02.5.2.5). The physical requirement that the two-phase mixture should have the incompressible properties of either the matrix or the fluid when $\phi = 0$ or $\phi = 1$ imposes a porosity dependence to ζ with $\lim_{\phi\to 0}\zeta(\phi) = \lim_{\phi\to 1}\zeta(\phi) = +\infty$. Simple micromechanical models (e.g., Bercovici et al., 2001a; Nye, 1953; Schmeling, 2000; Simpson et al., 2010b) allow us to estimate the bulk viscosity as

$$\zeta = K_0\frac{(\eta_f + \eta_m)}{\phi(1-\phi)} \qquad [234]$$

The dimensionless constant K_0 accounts for grain/pore geometry and is of $O(1)$.

A more general, but more hypothetical, interpretation of the entropy positivity could argue that some deformational work, Ψ, might affect the pressure drop of eqn [232]. This hypothesis led to a two-phase damage theory developed in Bercovici and Ricard (2003, 2005), Bercovici et al. (2001a,b), and Ricard and Bercovici (2003) (see also Bercovici et al., this volume). Here, we assume that the system remains close enough to mechanical equilibrium that damage does not occur.

The second equation [233] controls the kinetics of the melting/freezing and, by consequence, defines the equilibrium condition. In the case of mechanical equilibrium, when there is no pressure drop between the two phases, the melting rate cancels when there is equality of the chemical potentials of the two single phases. In case of mechanical disequilibrium, ($\Delta P \neq 0$), the chemical equilibrium does not occur when the two chemical potentials are equal. We define a new effective chemical potential,

$$\mu_i^* = \mathcal{U}_i + \frac{P^\omega}{\rho_i} - TS_i \qquad [235]$$

where i stands for f or m, and write the kinetic equation [233],

$$\Delta\Gamma = \mathcal{R}\Delta\mu^* \qquad [236]$$

Chemical equilibrium imposes the equality of the effective potentials on the interface, at the pressure P^ω at which the phase change effectively occurs.

Using eqns [232] and [233], we can show that the entropy production is indeed positive and given by

$$TH_S = k\frac{1}{T}|\nabla T|^2 + \frac{\Delta P^2}{\zeta} + \mathcal{R}(\Delta\mu^*)^2 + \Psi + \overline{\rho}\overline{H} \qquad [237]$$

Chemical relaxation and bulk compression are associated with entropy sources, in addition to diffusion, viscous dissipation, and radioactive heating (see also Rudge et al., 2011).

7.02.5.2.5 Summary equations

For convenience, we summarize the governing equations when the matrix is much more viscous than the fluid phase ($\eta_f \ll \eta_m$) as typical for melting scenarios, which implies that $\underline{\boldsymbol{\tau}}_f = 0$, $\omega = 0$, $\rho^\omega = \rho_f$, $P^\omega = P^\omega = P_f$, and $\mathbf{v}_\omega = \mathbf{v}_m$.

The mass conservation equations are [205] and [206]. The equation of conservation of momentum for the fluid phase is

$$-\phi\nabla[P_f + \rho_f gz] + c\Delta\mathbf{v} = 0 \qquad [238]$$

(assuming z positive upward). This is the Darcy law with $c = \eta_f\phi^2/k(\phi)$, where k is the permeability (often varying as $k_0\phi^n$ with $n = 2$ or 3 (Simpson et al., 2010b)). The second momentum equation could be the matrix momentum (eqn [212]) or a combined force-difference (or action–reaction) equation:

$$-\nabla[(1-\phi)\Delta P] + (1-\phi)\Delta\rho\mathbf{g} + \nabla\cdot\left[(1-\phi)\underline{\boldsymbol{\tau}}_m\right] - \frac{c\Delta\mathbf{v}}{\phi} = 0 \qquad [239]$$

where the deviatoric stress in the matrix is given by

$$\underline{\boldsymbol{\tau}}_m = \eta_m\left(\nabla\mathbf{v}_m + [\nabla\mathbf{v}_m]^T - \frac{2}{3}\nabla\cdot\mathbf{v}_m\underline{\mathbf{I}}\right) \qquad [240]$$

and the pressure jump between phases, eqn [232], becomes

$$\Delta P = -\zeta(1-\phi)\nabla\cdot\mathbf{v}_m \qquad [241]$$

if an equivalent bulk viscosity is used, or

$$\Delta P = -K_0\eta_m\frac{\nabla\cdot\mathbf{v}_m}{\phi} \qquad [242]$$

from the micromechanical model, eqn [234], of Bercovici et al. (2001a) or of Simpson et al. (2010b).

The action–reaction equation [239] can also be written in a different way, for example, by elimination of $\Delta\mathbf{v}$ taken from the Darcy equilibrium [238],

$$-\nabla P_f + \nabla\cdot\left[(1-\phi)\underline{\boldsymbol{\tau}}_m^*\right] + \overline{\rho}\mathbf{g} = 0 \qquad [243]$$

where $\underline{\boldsymbol{\tau}}_m^*$ includes the ΔP term and is defined by

$$\underline{\boldsymbol{\tau}}_m^* = \eta_m\left(\nabla\mathbf{v}_m + [\nabla\mathbf{v}_m]^T - \frac{2}{3}\nabla\cdot\mathbf{v}_m\mathbf{I}\right) + \zeta(1-\phi)\nabla\cdot\mathbf{v}_m\mathbf{I} \qquad [244]$$

This shows that if the pressure is defined everywhere as the fluid pressure, then it is equivalent to use for the matrix a rheology (see eqn [95]), with a bulk viscosity $(1-\phi)\zeta \approx \zeta$ (McKenzie, 1984). This analogy only holds without surface tension between phases (Bercovici and Ricard, 2003; Bercovici et al., 2001a; Ricard et al., 2001).

The rate of melting is controlled by

$$\Delta\Gamma = \mathcal{R}\left(\Delta\mathcal{U} + P_f\left(\frac{1}{\rho_m} - \frac{1}{\rho_f}\right) - T\Delta S\right) \qquad [245]$$

and the energy equation is

$$\begin{aligned}\rho_f\phi C_f\frac{D_f T}{Dt} + \rho_m(1-\phi)C_m\frac{D_m T}{Dt} - \Delta\mathcal{H}\Delta\Gamma \\ = \overline{\rho}\overline{H} - \nabla\cdot\mathbf{q} + \frac{\Delta\Gamma^2}{\mathcal{R}} + K_0\eta_m\frac{1-\phi}{\phi}(\nabla\cdot\mathbf{v}_m)^2 + \Psi\end{aligned} \qquad [246]$$

where we have assumed the relation [242].

These equations have been used by many authors with various levels of approximation (Simpson et al., 2010a). The most benign have been to replace $1-\phi$ by 1. Most authors have also considered the bulk viscosity ζ as a porosity-independent parameter (e.g., Choblet and Parmentier, 2001; Katz et al., 2004; Kelemen et al., 1997; McKenzie, 1984; Ribe, 1985a,b; Richter and Mckenzie, 1984; Scott and Stevenson, 1984, 1986, 1989; Spiegelman and Kelemen, 2003; Spiegelman et al., 2001; Stevenson, 1989). This overestimates the possibilities of matrix compaction at low porosity. Porosity-dependent parameters have been explicitly accounted for in other papers (e.g., Bercovici et al., 2001a; Connolly and Podladchikov, 1998; Fowler, 1985; Rabinowicz et al., 2002; Ricard et al., 2001; Schmeling, 2000; Sleep, 1988).

For direct application to magma extraction, an earlier work by Spiegelman and McKenzie (1987) imposed matrix motions and solved for the magma trajectories that were found attracted toward ridges and subduction corners. The melting rates were later sometimes imposed (e.g., Turcotte and Morgan, 1992), sometimes solved according to the energy equation, assuming univariant melting (Fowler, 1989; Hewitt and Fowler, 2008; Sramek et al., 2007) or even multiple component melting (Hewitt, 2010; Rudge et al., 2011). Surface tension (between matrix and magma and between matrix grains) has been considered (Bercovici and Ricard, 2003; Bercovici et al., 2001a;

Hier-Majumder et al., 2006; Riley et al., 1990). The situation where the melt is denser than the matrix has been considered by Karato et al. (2006) and by Hernlund and Jellinek (2010) for applications to the mantle transition zone and to the deep mantle, two layers where the presence of magma is suspected from seismological observations (Tauzin et al., 2010; Williams and Garnero, 1996). The interaction between the matrix and the magma can lead to the channelization of the melt (Hewitt and Fowler, 2009; Weatherley and Katz, 2012) and to the development of a fabric of the matrix that favors the migration of the melt to the ridge (Holtzman et al., 2003).

Similar equations have also been used to describe the interaction between iron and silicates near the CMB (Buffett et al., 2000) or during the segregation of planetary cores (Golabek et al., 2008; Ricard et al., 2009; Sramek et al., 2010, 2012), the formation of dendrites (Poirier, 1991), the compaction of lava flows (Massol et al., 2001; Michaut and Bercovici, 2009), and the dynamics in volcanic conduits (Bercovici and Michaut, 2010; Michaut et al., 2009, 2013).

7.02.6 Specifics of Earth's Mantle Convection

In this last section, we discuss various aspects of physics unique to large-scale mantle convection. We are aware of the impossibility to be exhaustive, but most of the important points are more deeply developed in other chapters of the treatise (in particular, we leave the problems of partial melting to Parmentier and to Ito and Van Keken, the connections with geochemistry to Tackley, and the aspects of plate tectonics to Bercovici et al., this volume; see also the books by Schubert et al. (2001) and Davies (1999)).

7.02.6.1 A Negligible Inertia

The most striking difference between mantle convection and most other forms of convection is that inertia is totally negligible. This is because the Prandtl number is much larger than the (already very large) Rayleigh number. This implies that the mantle velocity flow obeys

$$\begin{aligned} \boldsymbol{\nabla}\cdot(\overline{\rho}\mathbf{v}) &= 0 \\ -\boldsymbol{\nabla}P + \boldsymbol{\nabla}\cdot\underline{\boldsymbol{\tau}} + \delta\rho\mathbf{g} + \overline{\rho}\delta\mathbf{g} &= 0 \\ \nabla^2\psi &= 4\pi G\delta\rho \\ \delta\mathbf{g} &= -\boldsymbol{\nabla}\psi \end{aligned} \qquad [247]$$

in agreement with eqn [137]. In this set of equation, we keep the self-gravitation term as appropriate at long wavelengths. If the internal loads $\delta\rho$ are known, the flow can be computed independently of the temperature equation. This time-independent system has been used by many authors to infer the mantle flow properties.

7.02.6.1.1 Dynamic models

The system of eqn [247] can been solved analytically for a depth-dependent viscosity, when variables are expressed on the basis of spherical harmonics (see Hager and Clayton, 1989 and also Ribe, this volume and Forte, vol. 3). Various possible surface observables (geoid height or gravity free-air anomalies, velocity divergence, amplitude of deviatoric stress at the surface, surface dynamic topography, CMB topography, etc.) can be expressed on the basis of spherical harmonics with components O_{lm}. Through eqn [247], they are related to the spherical harmonics components of the internal density variations $\delta\rho_{lm}(r)$ by various degree-dependent Green functions (Green, 1793–1841):

$$O_{lm} = \int G_l^O(r)\delta\rho_{lm}(r)\,\mathrm{d}r \qquad [248]$$

(see Ribe, this volume, for analytic details). The Green functions $G_l^O(r)$ can be computed from the averaged density and viscosity profiles.

Before seismic imaging gave us a proxy of the 3-D density structure of the mantle, various theoretical attempts have tried to connect models of mantle convection to plate velocities (Hager and Oconnell, 1979; Hager and O'Connell, 1981), to Earth's gravity field (or to the geoid, proportional to the potential ψ), to the lithospheric stress regime, or to the topography (Bai et al., 1992; Lago and Rabinowicz, 1984; Parsons and Daly, 1983; Ricard et al., 1984; Richards and Hager, 1984; Runcorn, 1964).

An internal load of negative buoyancy induces a downwelling flow that deflects the Earth's surface, the CMB, and any other internal compositional boundaries, if they exist. The amount of deflection corresponds to the usual isostatic rule for a load close to an interface: the weight of the induced topography cancels to first order the weight of the internal load. The total gravity anomaly resulting from a given internal load is affected by the mass anomalies associated with the flow-induced boundary deflections as well as by the load itself. Due to the deflection of the Earth's surface, the geoid perturbation induced by a dense sinking anomaly is generally negative (e.g., free-air gravity has a minimum above a dense load). However, when the mantle viscosity increases significantly with depth, by 1–2 orders of magnitude, a mass anomaly close to the viscosity increase, induces a larger CMB deformation and a lower surface deformation. The resulting gravity anomaly corresponds to a geoid high. The fact that cold subduction zones correspond to a relative geoid high suggests a factor $\geq$30 viscosity increase around the upper–lower mantle interface (Hager et al., 1985; Lago and Rabinowicz, 1984; Ricard et al., 1984; Richards and Hager, 1984). Shallow anomalies and anomalies near the CMB, being locally compensated, do not contribute to the long-wavelength gravity field. The lithospheric stress field, like the geoid, is affected by mid-mantle density heterogeneities. The surface deflection induced by a deep-seated density anomaly decreases with the depth of this anomaly, but even lower mantle loads should significantly affect the surface topography.

7.02.6.1.2 Mantle flow and postglacial models

As soon as seismic tomography started to image the mantle structures, these seismic velocity anomalies have been used to further constrain the mantle viscosity. The fact that the geoid and seismic velocity anomalies are positively correlated around the transition zone but negatively in the deep mantle heterogeneities suggests a viscosity larger than 10 but not too large (less than 100); otherwise, the mantle would be everywhere

positively correlated with gravity (Hager and Clayton, 1989; Hager et al., 1985; King and Hager, 1994; Ricard et al., 1984; Richards and Hager, 1984) (see **Figure 6**). The same modeling approach, assuming a proportionality between seismic velocity anomalies and density variations, was also used to match the observed plate divergence (Forte and Peltier, 1987), the plate velocities (Ricard and Vigny, 1989; Ricard et al., 1991), and the lithospheric stresses (Bai et al., 1992; Lithgow-Bertelloni and Guynn, 2004). The initial Boussinesq models were extended to account for compressibility (Forte and Peltier, 1991). The instantaneous flow computations that were initially restricted to long-wavelength patterns have recently capitalized on advances in adaptative mesh refinements on parallel computers to reach very fine resolutions, down to kilometer scales (Alisic et al., 2012; Stadler et al., 2010).

Joint inversions of gravity with postglacial rebound were also performed to further constrain the mantle viscosity profile. The viscosity increase required by subduction was initially thought to be too large to reconcile with postglacial rebound (Peltier, 1996). The various approaches (time-dependent for the postglacial models and time-independent for the geoid models) seem to have converged to a standard viscosity profile with a significant increase with depth (Mitrovica and Forte, 1997). Whether this viscosity increase occurs across a discontinuity (at the upper–lower mantle interface or deeper) or as a gradual increase is probably beyond the resolution of these approaches.

Although these dynamic models explain the observed geoid, they require surface topography of order 1 km, induced by mantle convection and called dynamic topography (in contrast to the isostatic topography related to crustal and lithospheric density variations). The direct observation of dynamic topography, however, is not evident in Earth's topography corrected for isostatic crustal contributions (e.g., Colin and Fleitout, 1990; Kido and Seno, 1994; Lestunff and Ricard, 1995; Lithgow-Bertelloni and Silver, 1998). However, in recent years, the role of a deep dynamic support to explain the present-day topography and the flooding and emergence of continent interiors during the geologic times have been documented by various authors (Conrad and Gurnis, 2003; Conrad and Husson, 2009; Flament et al., 2013; Husson and Conrad, 2006; Moucha et al., 2008).

7.02.6.1.3 Time-dependent models

The thermal diffusion in the mantle is so slow that even over 100–200 My, it can be neglected in some long-wavelength global models. Equation [247] can thus be solved by imposing the known paleoplate velocities at the surface and advect the mass anomalies with the flow without explicitly solving the energy equation. This forced-convection approach has shown that the deep mantle structure is mostly inherited from the Cenozoic and Mesozoic plate motion (Lithgow-Bertelloni and Richards, 1998; Richards and Engebretson, 1992). From plate paleoslab reconstructions only, a density model can be obtained that gives a striking fit to the observed geoid and is in relative agreement with long-wavelength tomography (Ricard et al., 1989). This approach was also used to study the hot spot fixity (Richards, 1991; Steinberger and O'Connell, 1998), the sea level changes (Lithgow-Bertelloni and Gurnis, 1997), or the polar wander of the Earth (Richards et al., 1997; Spada et al., 1992a).

Instead of an imposed forced convection, a promising method proposes to infer self-coherently the past mantle structure from the whole set of convection equations, including temperature evolution (Bunge et al., 2003; Ismail-Zadeh et al., 2004; Liu and Gurnis, 2008). Using tools of variational data assimilation developed by meteorologists, the past density heterogeneities can be considered as unknowns to be inverted for in such a way that their time evolution is in agreement with present-day observations (e.g., with tomographic models), paleoreconstructions (the geometry and velocities of plates), or paleogeography (Shephard et al., 2010).

7.02.6.2 A Mantle with Internal Heating

When the top and bottom boundary conditions are the same (i.e., both free-slip or both no-slip), purely basally heated convection in a Cartesian box leads to a perfectly symmetrical system. We could simultaneously reverse the vertical axis and the color scale of **Figure 4** and get temperature patterns that are also convective solutions. The convective fluid has a near adiabatic core and the temperature variations are confined into two boundary layers, a hot bottom layer and a cold top layer. The thicknesses of these two boundary layers and the temperature drops across them are the same. The middepth temperature is simply the average of the top and bottom temperatures. Instabilities develop from the bottom layer (hot rising plumes) and the cold layer (cold downwelling plumes). They have a temperature hotter or colder than the depth-dependent average temperature. They are active structures driven by their intrinsic positive or negative buoyancy. The Earth's mantle has however a large number of characteristics that break the symmetry between upwellings and downwellings.

What is probably the major difference between mantle convection and purely basally heated convection is that the Earth is largely powered by radiogenic heating from the decay of uranium, thorium, and potassium and loss of primordial heat left over from its accretion. Convection purely heated from within is depicted in **Figure 7**. In the extreme case where the fluid is entirely heated from within, the fluid has no hot bottom boundary layer. There are only concentrated downwelling currents sinking from the top cold boundary layer. The downwellings are active as they are moved by their own negative buoyancy. To compensate for the resulting downwelling flow, the background rises passively, that is, without being pushed up by a positive buoyancy (Bercovici et al., 2000). In the case of basal heating, any plume leaving the top or bottom boundary layer travels adiabatically (neglecting diffusion and shear heating). However, in the case of internal heating, while the rapid downwellings remain close to adiabatic, the radioactive decay can accumulate heat within the slow upwellings. This heating is opposite to the adiabatic cooling and the average temperature in an internally heated system remains more homogeneous and with a significant subadiabatic gradient (Parmentier et al., 1994).

The Earth's mantle is however not in such an extreme situation. Some heat flow is extracted across the CMB from the molten iron outer core. This basal heat flux drives active upwellings (hot spots). The ratio of the internal radioactive heat to the total heat extracted at the Earth's surface is called the Urey number (Urey, 1951). Geochemical models of mantle composition (McDonough and Sun, 1995; Rudnick and Fountain, 1995) imply that about 50% of the surface heat flux is due to mantle and core cooling and only 50% or even

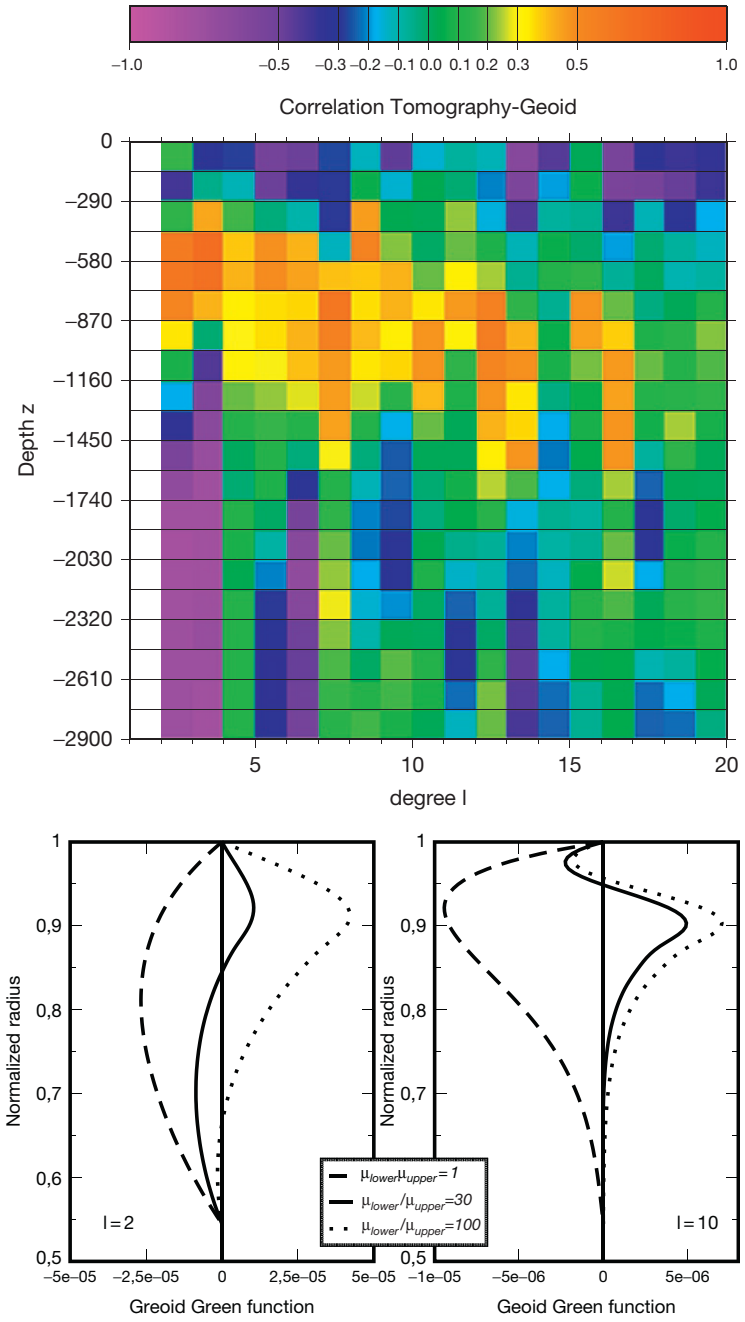


Figure 6 Correlations between gravity and the synthetic tomographic SMEAN model (Becker and Boschi, 2002) as a function of degree *l* and normalized radius (top). The seismic velocities, proxy of the density variations, are positively correlated with gravity around the upper–lower mantle interface (warm colors) but negatively correlated near the surface and in the deep lower mantle (cold colors). Geoid Green functions for degree 2 (bottom left) and degree 10 (bottom right) and three possible viscosity increases between upper mantle and lower mantle. The geoid Green function for a uniform viscosity (dashed line) is everywhere negative, and all the anomalies around the upper–lower mantle would induce a gravity signal opposite to that observed. A too large viscosity increase (a factor 100 for the dotted lines) cannot explain the rather good negative correlation between lower mantle anomalies and the geoid at long wavelength. A moderate increase (a factor 30 for the solid line) leads to the best fit as the sign of the Green functions is everywhere that of the observed density–gravity correlations. The different Green functions are computed for an incompressible mantle, with a lithosphere, ten times more viscous than the upper mantle.

less (Lyubetskaya and Korenaga, 2007) to radioactive decay. Generally, geophysicists have difficulties with these numbers as they seem to imply a too large mantle temperature in the past (Davies, 1980; Schubert et al., 1980). From convection modeling of the Earth's secular cooling, they often favor ratios of the order of 80% radioactive and 20% cooling. A more complex coupling between core and mantle cooling (Labrosse et al., 2007), or the peculiar properties of the lithosphere, may reconcile the thermal history of the Earth with a low radiogenic content (Grigne et al., 2005; Korenaga, 2003) (see Jaupart et al. and Hernlund and McNamara, this volume). The basal heat flux at the CMB represents the core cooling component that is

part of the total cooling rate of the Earth. The secular cooling and the presence of internal sources tend to decrease the thickness of the hot bottom layer compared to that of the cold top layer, increase the active role of downwellings (the subducting slabs), and decrease the number or the strength of the active upwellings (the hot spots).

7.02.6.3 A Complex Rheology

We have shown that the rheological laws of crystalline solids may be linear or nonlinear, depending on temperature, grain size, and stress level. Various deformation mechanisms (grain diffusion, grain boundary diffusion, dislocation creep, grain boundary sliding, etc.) act simultaneously. The equivalent viscosity of each individual mechanism can be written in the form

$$\eta = AI_2^{-\frac{(n-1)}{2}} d^{-m} \exp\frac{E^* + PV^*}{RT} \qquad [249]$$

where E^* and V^* are the activation energy and volume, P and T the pressure and temperature, R the perfect gas constant, d the grain size, m the grain size exponent, I_2 the second stress invariant, and n a stress exponent (Ranalli, 1995; Weertman and Weertman, 1975). The multiplicative factor A varies with water content, melt content, and mineralogy. In general, the composite rheology is dominated by the mechanism leading to the lowest viscosity.

In Figure 8, we plot as a function of temperature, and for various possible grain sizes (0.1 mm, 1 mm, and 1 cm), the stress rate at which the strain rate predicted for the dislocation mechanism and that for diffusion mechanism are the same (see eqns [108] and [107]). The data correspond to dry upper mantle (Karato and Wu, 1993). Low stress and temperature favor diffusion creep, while high stress and high temperature favor dislocation creep. These experiments suggest that in the upper mantle or at least in its shallowest part, nonlinear creep is likely to occur. At greater depth, the decrease in the average deviatoric stress should favor a diffusive regime with a Newtonian viscosity. The observation of rheological parameters at lower mantle conditions is more difficult, but the lower mantle should mostly be in diffusive creep regime except the zones of intense shear around subductions (McNamara et al., 2001).

The lateral variations of viscosity due to each separate parameter, stress exponent, temperature, water content, or grain size can potentially be very large. Surprisingly, attempts to deduce these variations directly from geodynamic observations have not been very successful. Attempts to explain the Earth's gravity from internal loads do not seem to require lateral viscosity variations in the deep mantle (Zhang and

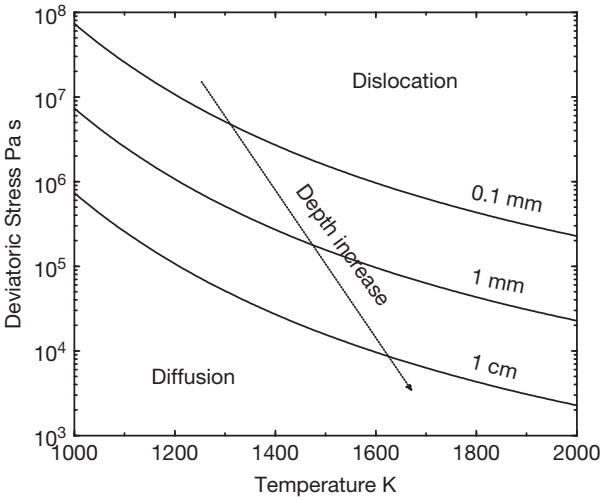


Figure 8 Creep regime map for dry olivine. High deviatoric stress or temperature favors a dislocation mechanism. A decrease in the grain size favors diffusion. In the upper mantle, the stress and temperature conditions tend to bring the creep regime from dislocation to diffusion at depth ($n = 2.5$, $m = 0$, and $E^* = 540$ kJ mol^{-1} for dislocation creep and $n = 0$, $m = 2.5$, and $E^* = 300$ kJ mol^{-1} for diffusion creep).

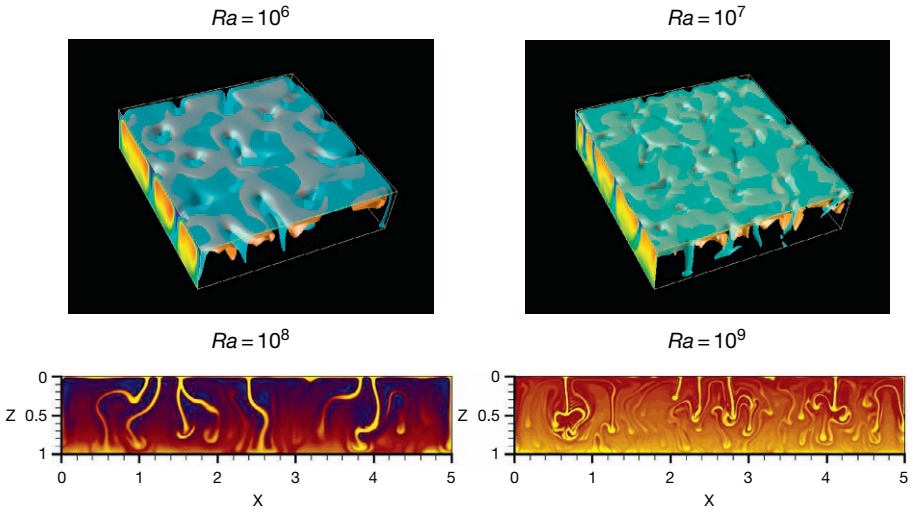


Figure 7 Convection patterns of a fluid entirely heated from inside at Rayleigh number 10^6, 10^7, 10^8, and 10^9 (simulation compliments of Fabien Dubuffet). Cold fingerlike instabilities are sinking from the top boundary layer and spread on the bottom boundary layer. No active upwellings are present (compare with convection patterns for a fluid heated from below, Figure 4).

Christensen, 1993). Near the surface, viscosity variations are present, at least between continental and oceanic lithospheres (Cadek and Fleitout, 2003; Ricard et al., 1991). The gain in fitting the Earth's gravity or postglacial rebound data with 3-D viscosity models remains rather moderate compared to the complexities added to the modeling (Gasperini et al., 2004), and most models of mantle viscosity are restricted to radial profiles (Mitrovica and Forte, 1997). Even the modeling of slabs, their curvatures, and their stress patterns do not really require that they are much stiffer than the surrounding mantle (Tao and Oconnell, 1993).

7.02.6.3.1 Temperature dependence of viscosity

Mantle viscosity is a strong function of temperature and the cold lithosphere seems to have a viscosity of order 10^{25} Pa s (Beaumont, 1978), 4–6 orders of magnitude stiffer than the asthenosphere. The activation energy E^* is typically from 300 to 600 kJ mol^{-1} (Drury and FitzGerald, 1998) the lowest values being for diffusion creep. This implies a factor ≈ 10 in viscosity decrease for a 100 K temperature increase (using $T \approx 1600$ K). The effect of temperature dependence of viscosity on the planform of convection was recognized experimentally using oils or syrups (Booker, 1976; Nataf and Richter, 1982; Richter, 1978; Weinstein and Christensen, 1991). In the case of a strongly temperature-dependent viscosity, the definition of the Rayleigh number is rather arbitrary as the maximum, the minimum, or some average viscosity can be chosen in its definition. Another nondimensional number (e.g., the ratio of viscosity variations, η_{max}/η_{min}) must be known to characterize the convection.

Not surprisingly, two extreme regimes are found. For a viscosity ratio lower than about 100, the convection pattern remains quite similar to convection with uniform viscosity. On the other hand, if the viscosity of the cold boundary layer (the lithosphere) is more than 3000 times that of the underlying asthenosphere, the surface becomes stagnant (Solomatov, 1995). Below this immobile lid, the flow resembles convection below a rigid top surface (Davaille and Jaupart, 1993). In between, when the viscosity ratios are in the range 100–3000, the lithosphere deforms slowly, and in this sluggish regime, the convection cells have large aspect ratios.

Convection with temperature-dependent viscosity has been investigated by various authors (Christensen, 1984b; Kameyama and Ogawa, 2000; Parmentier et al., 1976; Tackley et al., 1993; Trompert and Hansen, 1998b). Since the top boundary layer is stiffer than the bottom boundary layer, the top boundary layer is also thicker than the bottom one. This impedes surface heat removal, reduces the heat flux across the bottom boundary layer, and raises the average mantle temperature. Convection patterns computed with T-dependent viscosity remain however quite far from Earth-like convection. The major difference is that when the T dependence is too strong, the surface freezes and becomes immobile, while on the real Earth, the lithosphere is highly viscous but broken into tectonic plates separated by weak boundaries. Without mechanisms other than a simple T dependence of viscosity, the Earth would be in a stagnant-lid regime (see Bercovici et al., this volume).

Various modelers have thus tried to use T-dependent rheologies but have imposed a platelike surface velocity. This has been very useful to understand the initiation of subduction (Toth and Gurnis, 1998), the interaction of slabs with the phase changes in the transition zone (Christensen, 1996, 1997b), and the relationship between subduction and gravity (King and Hager, 1994). These numerical experiments, mostly intended to model slabs, compare satisfactorily with laboratory experiments (Guilloufrottier et al., 1995; Kincaid and Olson, 1987).

To conclude this brief section on temperature dependence of viscosity, we discuss the general concept of self-regulation of planetary interiors (Tozer, 1972). If a planet were convecting too vigorously, it would lose more heat than radioactively produced. It would therefore cool down until the viscosity is large enough to reduce the heat transfer. In contrast, a planet convecting too slowly would not extract its radioactive energy and would heat up until the viscosity is reduced sufficiently to increase the surface heat flow (see also Jaupart et al. this volume). The internal temperature of planets is mostly controlled to the activation energy (or rather enthalpy) of the viscosity (assuming that planets have similar amount of heat sources). To first order, large and small terrestrial rocky planets probably have the same internal temperatures.

7.02.6.3.2 Depth dependence of viscosity

The activation volume V^* of the viscosity [249] is typically around 10^{-5} m^3 mol^{-1}. Extrapolating to CMB conditions, this suggests a large viscosity increase throughout the mantle. However, measurements of viscosity at both high T and P conditions are very difficult (see also Weidner, vol. 2). The viscosity increase by a factor 30–100 suggested by geodynamics (see **Section 7.02.6.1**) is probably as robust a constraint as could be deduced from mineral physics experiments.

The effect of a depth-dependent viscosity on the planform of convection has been studied by, for example, Gurnis and Davies (1986), Cserepes (1993), and Dubuffet et al. (2000). At least two important geodynamic observations can be explained by an increase of viscosity with depth. One is the relative stability of hot spots. A sluggish lower mantle where convection is decreased in intensity by a larger viscosity (and also by a smaller expansivity and a potentially larger thermal conductivity as discussed in **Section 7.02.6.5.3**) favors the relative hot spot fixity (Richards, 1991; Steinberger and O'Connell, 1998). A second consequence is a depth dependence of the wavelengths of the thermal heterogeneities. A viscosity increase (together with the existence of plates and continents that impose their own wavelengths (see **Section 7.02.6.7**)) induces the existence of large-scale thermal anomalies at depth (Bunge and Richards, 1996). A slab crossing a factor 30–100 viscosity increase should thicken by a factor of order 3–5 (Gurnis and Hager, 1988; Ribe et al., 2007). This thickening is observed in tomographic models (van der Hilst et al., 1997) and can be inferred from geoid modeling (Ricard et al., 1993a). When the viscosity increases with pressure, the self-regulation effect causes the mantle to heat up. The temperature dependence of the viscosity then mitigates the effects of the pressure dependence (Tackley et al., 2013).

7.02.6.3.3 Stress dependence of viscosity

Starting from Parmentier et al. (1976), the effect of a stress-dependent viscosity has been studied by Christensen (1984a), Malevsky and Yuen (1992), van Keken et al. (1992), and Larsen

et al. (1993), assuming either entirely nonlinear or composite rheologies (where deformation is accommodated by both linear and nonlinear mechanisms). At moderate Rayleigh number, the effect of a nonlinear rheology is not very significant. In fact, the nonlinearity in the rheology is somewhat opposed to the temperature dependence of the rheology. As shown by Christensen (1984a), a T-dependent, nonlinear rheology with an exponent $n \approx 3$ leads to convection cells rather similar to what would be obtained with a linear rheology and an activation energy divided by $\approx n$. Convection with both nonlinear and T-dependent rheology looks more isoviscous than convection with only stress-dependent or only T-dependent rheologies.

At large Rayleigh number, however, nonlinear convection becomes more unstable (Malevsky and Yuen, 1992), and the combination of nonlinear rheology, T-dependent rheology, and viscous dissipation can accelerate the rising velocity of hot plumes by more than an order of magnitude (Larsen and Yuen, 1997).

7.02.6.3.4 Grain size dependence of viscosity

In the diffusion regime, the viscosity [249] can be strongly variable as controlled by the grain size d to a power m of order 3 (Karato et al., 1986). The difficulty to estimate the rheological conditions of the mantle is that the grain size d of eqn [249] has its own dynamics with various feedbacks with the deformation itself. In the diffusive regime, the grains are slowly coarsening: the slight overpressure of small grains due to their surface tension drives a diffusional flow toward the larger grains (e.g., Hillert, 1965). In contrast, grains undergoing dislocation creep can go through recrystallization and therefore can reduce in size. These two opposite mechanisms should drive the mantle material in a situation where diffusion and dislocation creeps have roughly the same efficiency (De Bresser et al., 2001; Rozel et al., 2011). The existence of an equilibrium grain size related to the average deviatoric stress is observed in laboratory experiments (Karato, 2008; Vanderwal et al., 1993) and inferred from thermodynamic considerations (Austin and Evans, 2007; Bercovici and Ricard, 2005; Ricard and Bercovici, 2009).

There is a potential feedback interaction between deformation, grain size reduction by dynamic recrystallization, viscosity reduction, and further localization (Braun et al., 1999; Jaroslow et al., 1996; Kameyama et al., 1997). However, in its simplest version, this localizing mechanism is problematic as it implies simultaneously dynamic recrystallization and diffusion creep that ostensibly occur in exclusive domains of deformation space. It has however been proposed that the presence of minor mineralogical phases can both hinder the grain coarsening (Hiraga et al., 2010) and ease the recrystallization, thus providing a consistent mechanism for localization (Bercovici and Ricard, 2012, 2014; Warren and Hirth, 2006). The effect is potentially important in the mantle and even more important in the lithosphere. More details are given in Bercovici et al. this volume.

7.02.6.4 Importance of Sphericity

An obvious difference between the convection planform in a planet and that of in an experimental tank is due to the sphericity of the former. In the case of purely basally heated convection, the same heat flux (in a statistical sense) has to be transported through the bottom boundary layer and the top boundary layer. However, as the CMB surface is about four times smaller than the top surface, this implies a four times larger thermal gradient through the bottom boundary layer than across the lithosphere. A large gradient across the bottom boundary layer reinforces the upwelling hot instabilities with respect to the downgoing cold instabilities. Sphericity affects the average temperature and the top and bottom boundary layer thicknesses in a way totally opposite to the effects of internal sources (see **Section 7.02.6.2**) or T-dependent viscosity (see **Section 7.02.6.3**). Although numerically more time-consuming, spherical convection models are more and more common (Bercovici et al., 1989a,b, 1992; Bunge et al., 1997; Choblet et al., 2007; Glatzmaier, 1988; Tackley et al., 1993; Zhong et al., 2000). In the last years, they reached a spectacular resolution (down to length scales of 1 km (Alisic et al., 2012; Stadler et al., 2010)) and robustness (handling up to 19 orders of magnitude variations in viscosity (Tackley, 2008)).

7.02.6.5 Other Depth-Dependent Parameters

7.02.6.5.1 Thermal expansivity variations

The thermal expansivity varies with depth, as predicted by the EoS [80], from which we can easily deduce that

$$\alpha = \frac{\alpha_0}{(\rho/\rho_0)^{n-1+q} + \alpha_0 (T - T_0)} \qquad [250]$$

where $n-1+q \simeq 3$.

It decreases with both temperature and density and thus with depth. The expansivity varies from $\approx 4 \times 10^{-5}\ \mathrm{K}^{-1}$ near the surface to $\approx 8 \times 10^{-6}\ \mathrm{K}^{-1}$ near the CMB (Chopelas and Boehler, 1992). This diminishes the buoyancy forces and slows down deep mantle convection (Hansen et al., 1993). Like the increase of viscosity with depth, a depth-dependent thermal expansivity broadens the thermal structures of the lower mantle and suppresses some hot instabilities at the CMB. On the other hand, hot instabilities gain buoyancy as they rise in the mantle, which favors their relative lateral stationarity. In addition to its average depth dependence, the temperature dependence of the expansivity also affects the buoyancy of slabs (Schmeling et al., 2003).

7.02.6.5.2 Increase in average density with depth

To take into account compressibility and the depth dependence of density, the Boussinesq approximation has been replaced by the anelastic approximation in several studies. Such investigations have been carried out by Jarvis and Mckenzie (1980), Glatzmaier (1988), and Bercovici et al. (1992) and since extended to higher Rayleigh numbers (e.g., Balachandar et al., 1992, 1993; Zhang and Yuen, 1996).

One of the difficulties with compressible fluids is that the local criterion for instability (see **Section 7.02.4.2.3**) is related to the adiabatic gradient. Depending on assumptions about the curvature of the reference geotherm with depth (the slope of the adiabatic gradient), part of the fluid can be unstable while the other part is stable. Assuming a uniform adiabatic gradient does not favor the preferential destabilization of either the upper or the lower mantle. On the other hand, assuming that the reference temperature increases exponentially with depth (i.e., taking eqn [123] with constant

parameters) would lead to an easier destabilization of the top of the mantle than of its bottom as a much larger heat flux would be carried along the lower mantle adiabat. In the real Earth, the adiabatic gradient (in K km^{-1}) should decrease with depth (due to the decrease in expansivity $\overline{\alpha}$ with depth insufficiently balanced by the density increase; see eqn [123]). Since less heat can be carried out along the deep mantle adiabat, compressibility should favor the destabilization of the deep mantle.

Compressible convection models generally predict downgoing sheets and cylindrical upwellings reminiscent of slabs and hot spots (e.g., Zhang and Yuen, 1996). Viscous dissipation is positive (as an entropy-related source) but maximum just below the cold boundary layer and just above the hot boundary layer, where rising or sinking instabilities interact with the layered structures of the boundary layers. On the contrary, the adiabatic source heats the downwellings and cools the upwellings. On average, it reaches a maximum absolute value in the midmantle. Locally, viscous dissipation and adiabatic heatings can be larger than radiogenic heat production, although integrated other the whole mantle and averaged over time, the adiabatic and dissipative sources cancel out (see eqn [61]).

7.02.6.5.3 Thermal conductivity variations

The thermal conductivity of a solid is due to two different effects. First, a hot material produces blackbody radiation that can be absorbed by neighboring atoms. This radiative transport of heat is probably a minor component since the mean free path of photons in mantle materials is very small. Second, phonons, which are collective vibrations of atoms, are excited and can dissipate their energies by interacting with other phonons, with defects and with grain boundaries. The free paths of phonons being larger, they are the main contributors to the thermal conductivity.

According to Hofmeister (1999), thermal conductivity should increase with depth by a factor $\approx$2–3. The recent observations of phase transitions in the bottom of the lower mantle should also be associated with another conductivity increase (Badro et al., 2004); this is one more effect (with the viscosity increase and the thermal expansivity decrease) that should decrease the deep mantle convective vigor. Thermal conductivity increase with depth also broadens the thermal anomalies and thins the bottom boundary layer (Dubuffet et al., 1999).

7.02.6.6 Thermochemical Convection

7.02.6.6.1 Density variations in the mantle

Except in **Section 7.02.5**, a simple negative relationship was assumed between density variations and temperature variations, through the thermal expansivity, $\Delta\rho = -\alpha\rho\Delta T$. However, in the mantle, several sources of density anomalies are present (see also, Stixrude, vol. 1). The density in the mantle varies with the temperature T for a given mineralogical composition, or phase content, symbolized by the symbol ϕ (e.g., for a given proportion of oxides and perovskite in the lower mantle). The mineralogy for a given bulk elemental composition χ (e.g., the proportion of Mg, Fe, and O atoms) evolves with pressure and temperature to maintain the Gibbs energy minimum. The variations of density in the mantle at a given pressure have potentially three contributions that can be summarized as

$$\Delta\rho = \left(\frac{\partial\rho}{\partial T}\right)_{\phi}\Delta T + \left(\frac{\partial\rho}{\partial\phi}\right)_{T}\left(\frac{\partial\phi}{\partial T}\right)_{\chi}\Delta T + \left(\frac{\partial\rho}{\partial\phi}\right)_{T}\left(\frac{\partial\phi}{\partial\chi}\right)_{T}\Delta\chi \quad [251]$$

The first term on the right side is the intrinsic thermal effect computed assuming a fixed mineralogy; we have already discussed this term. The second term is a thermochemical effect. The density is a function of the mineralogical composition controlled at uniform pressure and elemental composition, by the temperature variations. This effect is responsible for a rise in the 410 km deep interface and it deepens the 660 km interface in the presence of cold downwellings (Irifune and Ringwood, 1987). The last term is the intrinsic chemical effect (related to variations of the mineralogy due to changes in the elemental composition at constant temperature). The three contributions have very similar amplitudes and none of them is negligible (Ricard et al., 2005).

7.02.6.6.2 Phase changes

The effect of the second term has been rather well studied (Christensen, 1996; Christensen and Yuen, 1984; Machetel and Weber, 1991; Peltier and Solheim, 1992; Schubert et al., 1975; Tackley, 1995; Tackley et al., 1993). Phase changes in cold downgoing slabs occur at shallower depth in the case of exothermic phase changes and at greater depth for endothermic phase changes (the ringwoodite to oxides plus perovskite phase change at 660 km depth is endothermic; all the important other phase changes of the transition zone are exothermic). These sources of anomalies and their signs are related to the Clapeyron slope of the phase transitions (Schubert et al., 1975). The existence of latent heat release during phase change (see eqn [165]) is a secondary and minor effect. The recent discovery of a phase transformation in the deep lower mantle (Murakami et al., 2004) (the postperovskite phase) suggests that part of the complexities of the D″ layer is related to the interaction between a phase change and the hot boundary layer of the mantle (Nakagawa and Tackley, 2006) (see also Irifune, vol. 2).

The fact that below the normal 660 km depth interface there is a region where slabs remain in a low-density upper mantle phase instead of being transformed into the dense lower mantle phase is potentially a strong impediment to slab penetration. The idea that this effect induces a layering of convection at 660 km or a situation where layered convection is punctuated by large 'avalanche' events dates back to Ringwood and Irifune (1988) and was supported by numerical simulations in the 1990s (e.g., Honda et al., 1993; Machetel and Weber, 1991; Tackley, 1995). It seems however that the importance of this potential effect has been reduced in recent simulations with more realistic Clapeyron slopes, phase diagrams (taking into account both the pyroxene and the garnet phases), thermodynamic reference values (the phase change effect has to be compared with thermal effects and thus an accurate choice for the thermal expansivity is necessary), and viscosity profiles.

7.02.6.6.3 Abyssal layers

In the last years, various complexities were discovered in the deep lower mantle that was previously considered as rather

homogeneous. At small scale, a laterally intermittent layer at the base of D″ (ULVZ, ultralow-velocity zone), with a maximum thickness near 40 km and a strong decrease of the compressional wave velocity, is most simply explained as the result of partial melt at this depth (Williams and Garnero, 1996). A pair of seismic discontinuities observed in some fast (cold) regions of D″ could be the result of a double-crossing of the postperovskite phase boundary by the geotherm at two different depths (Hernlund et al., 2005). Two deep slow velocity anomalies under West Pacific and Africa (roughly underneath the two maxima of the geoid) have unusual seismic properties. They have an anomalously large ratio of compressional to shear velocity ratio, v_p/v_s (Masters et al., 2000), and an anticorrelation between density and seismic velocities (Ishii and Tromp, 1999) and between bulk and shear velocities (Kennett and Gorbatov, 2004) (the bulk velocity is defined as $\sqrt{K_s/\rho} = \sqrt{v_p^2 - (4/3)v_s^2}$). These anomalous regions have very sharp boundaries (Ni et al., 2002) and depending on the authors have been named megaplumes, thermochemical piles, or LLSVP (large low-shear-velocity province, according to Garnero and McNamara (2008)). Notice, however, that as these LLSVPs only cover part of the CMB surface, which is itself four times smaller than the Earth's surface, and as they only extend up to a few hundred kilometers, their total volume is not so large but only of order three times larger that of the continental crustal volume (Hernlund and Houser, 2008).

These observations of the deep mantle heterogeneity cannot easily be explained by temperature variations. They seem to require lateral variations of Fe or Si contents in the mantle (Bolton and Masters, 2001; Saltzer et al., 2004), although some authors interpret these observations in the framework of pure thermal models (e.g., Davies et al., 2012). The LLSVP should be intrinsically denser to resist entrainment by convection. The stability of deep dense layers has been discussed by various authors (Davaille, 1999; Kellogg et al., 1999; LeBars and Davaille, 2002; Samuel and Farnetani, 2003; Tackley and Xie, 2002). These compositional pyramids may anchor the hot spots (Davaille et al., 2002; Jellinek and Manga, 2002). The presence of a petrologically dense component of the source of hot spots also seems necessary to explain their excess temperature (Farnetani, 1997). These abyssal heterogeneities help to bridge the gap between geochemical observations and convection modeling (Coltice and Ricard, 1999; Coltice et al., 2011; van Keken et al., 2002).

Deep mantle heterogeneities may be a consequence of plate tectonics or be relics of the early evolution of our planet. A large well-documented elemental differentiation is between the oceanic crust (poor in Mg and rich in Al and Si) and the mantle. The oceanic crust in its high pressure eclogitic facies is ≈5% denser that the average mantle density in most of the mantle except in the shallowest 100 km of the lower mantle where it is lighter (Irifune and Ringwood, 1993).

In the deepest mantle, it is not yet totally clear whether the eclogite remains denser, neutrally buoyant, or even slightly lighter than the average mantle (e.g., Ricolleau et al., 2004). Thermochemical simulations starting with the pioneering paper of Christensen and Hofmann (1994) show the possibility of a partial segregation of oceanic crust during subduction, forming pyramidal piles on the CMB. These results have been confirmed (e.g., Davies, 2002; Nakagawa and Tackley, 2006; Tackley, 2000b). Not only present-day subductions can generate compositional anomalies in the mantle. Geochemists have often argued for a deep layer of primitive material. A possible origin of the LLSVP could be related to the formation of the core in the Hadean time. The presence of patches of melt on top of the CMB (Williams and Garnero, 1996) suggests that in the past, a much thicker deep magma ocean may have existed in between a deep solid mantle and the liquid core (Labrosse et al., 2007). The fractionation during the freezing of this magma could have resulted in a progressive enrichment of the residue (the LLSVP) and of the melt (the ULVZ) in dense components. This model therefore proposes a scenario for the formation of petrologic anomalies in the deep mantle (Coltice et al., 2011). This subject of mantle dynamics near the CMB is covered in greater detail by Hernlund and McNamara, this volume.

7.02.6.7 A Complex Lithosphere: Plates and Continents

The lithosphere is part of the convective circulation and plate tectonics and mantle convection cannot be separated. The fact that the cold lithosphere is much more viscous and concentrates most of the mass heterogeneities of the mantle makes it behaving to some extent like a membrane on top of a less viscous fluid. This suggests some analogy between mantle convection and what is called Marangoni convection (Marangoni, 1840–1925). Marangoni convection (Nield, 1964) is controlled by temperature-dependent surface tension on top of thin layers of fluids.

The Earth's mantle is certainly not controlled by surface tension and Marangoni convection, strictly speaking, has nothing to do with mantle convection. However, the equations of thermal convection with internal heating and with a highly viscous lithosphere can be shown to be mathematically related (through a change of variables) to those of Marangoni convection (Lemery et al., 2000). There are large differences between mantle convection and surface-driven convection, for example, because the presence of subducting slabs throughout the mantle affects the flow differently than what surface tractions can do. However, this analogy has sometimes been advocated as a 'top-down' alternative to mantle convection (Anderson, 2001), which is really not that different from what has been known for 30 years using boundary layer theory. Classically, the interpretation of plate cooling in terms of ridge-push force (Turcotte and Schubert, 1982) or the analysis of tectonic stresses using thin sheet approximations (England and Mckenzie, 1982) belongs to the same top-down approach.

Due to the complexities of the lithosphere properties, the boundary condition at the surface of the Earth is far from being a uniform free-slip condition. Both continents and tectonic plates impose their own wavelengths and specific boundary conditions on the underlying convecting asthenosphere. Of course, the position of the continents and the number and shape of the plates are themselves consequences of mantle convection. The plates obviously organize the large-scale flow in the mantle (Hager and Oconnell, 1979; Ricard and Vigny, 1989). They impose a complex boundary condition where the angular velocity is piecewise-constant. The continents with their reduced heat flow (Jaupart and Mareschal, 1999) also impose a laterally variable heat flux boundary condition.

Convection models with continents have been studied numerically (Grigné and Labrosse, 2001; Gurnis and Hager, 1988) and experimentally (Guillou and Jaupart, 1995). Continents with their thick lithosphere tend to increase the thickness of the top boundary layer and the temperature below them (see **Figure 9**). Hot rising currents are predicted under continents and downwellings are localized along continental edges. The existence of a thick and stable continental root must be due to a chemically lighter and more viscous subcontinental lithosphere (Doin et al., 1997). The ratio of the heat flux extracted across continents compared to that extracted across oceans increases with the Rayleigh number. This suggests that the continental geotherms were not much different in the past when the radiogenic sources were larger; it is mostly the oceanic heat flux that was larger (Lenardic, 1998). Simulating organized plates self-consistently coupled with a convective mantle has been a very difficult quest. The attempts to generate plates using T-dependent or simple nonlinear rheologies have failed. Although in 2-D some successes can be obtained in localizing deformation in platelike domains (Schmeling and Jacoby, 1981; Weinstein, 1996; Weinstein and Olson, 1992), they are obtained with stress exponents (e.g., $n \geq 7$) that are larger than what can be expected from laboratory experiments ($n \approx 3$). The problems are however worst in 3-D. Generally, these early models do not predict the important shear motions between plates that we observe (Christensen and Harder, 1991; Ogawa et al., 1991).

Some authors have tried to mimic the presence of plates by imposing platelike surface boundary conditions. These studies have been performed in 2-D and 3-D (Gable et al., 1991; King et al., 1992; Monnereau and Quéré, 2001; Ricard and Vigny, 1989). Although they have confirmed the profound effect of plates on the wavelengths of convection, on its time dependence and on the surface heat flux, these approaches cannot predict the evolution of surface plate geometry. **Figure 10** illustrates the organizing effect of plates in spherical, internally heated compressible convection with depth-dependent viscosity (Bunge and Richards, 1996). To obtain a self-consistent generation of surface plates, more complex rheologies that include brittle failure, strain softening, and damage mechanisms must be introduced (e.g., Auth et al., 2003; Bercovici, 1993, 1995; Moresi and Solomatov, 1998). The existence of plates seems also to require the existence of a weak sublithospheric asthenosphere (Richards et al., 2001). In the last ten years, the first successes in computing 3-D models that spontaneously organize their top boundary layer into plates have been reached using viscoplastic rheologies (Foley and Becker, 2009; Stein et al., 2004; Tackley, 1998, 2000c,d,e; Trompert and Hansen, 1998a; van Heck and Tackley, 2008). This important breakthrough made by the modelers allows convection simulations that display increasingly realistic seafloor spreading and continental drift, with plates, continents, ridges and one-sided subductions (Coltice et al., 2012; Rolf and Tackley, 2011; Rolf et al., 2012) (see **Figure 11**).

The Earth's plate boundaries keep the memory of their weakness over geologic times (Gurnis et al., 2000). This implies that the rheological properties cannot be a simple time-independent function of stress or temperature but have a long-term memory. The rheologies that have been used to predict plates in convective models remain empirical and their interpretation in terms of microscopic behavior and damage theory remains largely to be done (Bercovici and Ricard, 2005,

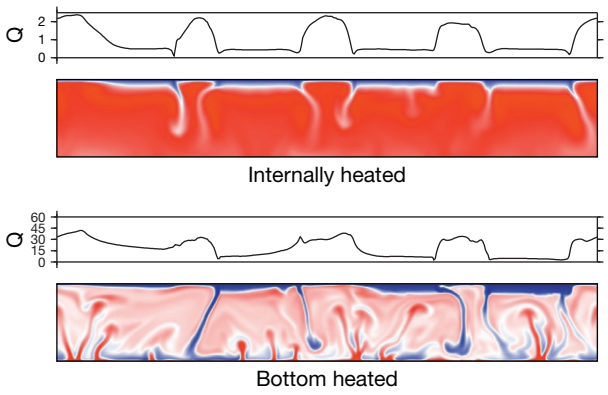


Figure 9 Convection patterns in the presence of four continents. The total aspect ratio is 7; the continents are defined by a viscosity increase by a factor 10^3 over the depth 1/10. The viscosity is otherwise constant. The Rayleigh number based on the total temperature drop (bottom panels) or on the internal radioactivity (top panels) is 10^7. The downwellings are localized near the continent margins. A large difference in heat flux is predicted between oceans and continents. In the case of bottom heating, hot spots tend to be preferentially anchored below continents where they bring an excess heat. This tends to reduce the surface heat flux variations.

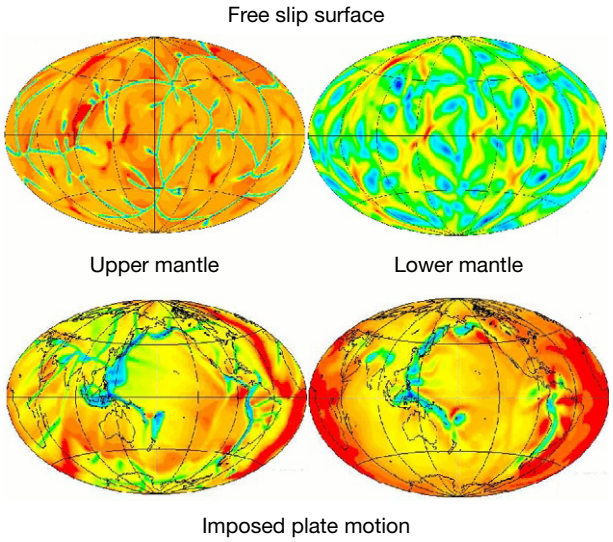


Figure 10 This figure depicts spherical compressible internally heated convection models where the viscosity increases with depth (simulations by Peter Bunge). In the first row, a uniform free-slip condition on top has been used. In the second row, the present-day observed plate motion is imposed at the surface. The left column shows the temperature field in the middle of the upper mantle and the right column in the middle of the lower mantle. This figure summarizes various points discussed in the text: the presence of linear cold downwellings, the absence of active upwellings in the absence of basal heating, and the enlargement of thermal structure in the more viscous lower mantle (top row). Although the modeling is not self-consistent (i.e., the presence of plates and the constancy of plate velocities are totally arbitrary), it is clear that the presence of plates can change radically the convection patterns (compare top and bottom rows).

2012, 2013, 2014). Reviews on the rapid progress and the limitations of self-coherent convection models can be found in Bercovici et al. (2000), Tackley (2000a), and Bercovici (2003) and are discussed in Bercovici et al., this volume.

7.02.6.8 Super-Earths

The study of the other planets of the solar system and their differences and similarities with the Earth is a very active research field described in the vol. 10 of the treatise. To the four terrestrial planets and the various dwarf planets and satellites of the solar system, many terrestrial extrasolar planets have been added in the last decade (e.g., Valencia et al., 2007b). A subset of these planets, approximately one to ten times as massive as Earth, have been termed super-Earths, because they are presumably analogous to our own planet. One can speculate that plate tectonics is a favorable condition for the existence of life. Plate tectonics buffers the climate by controlling the atmospheric carbon dioxide level with erosion, weathering, and subduction (Walker et al., 1981). Plate tectonics may also be necessary for the existence of life by providing a source of thermodynamic disequilibrium through continuous recycling of the surface (e.g., Southam and Westall, vol. 10). The question of whether super-Earths are more or less favorable to plate tectonics has therefore been debated.

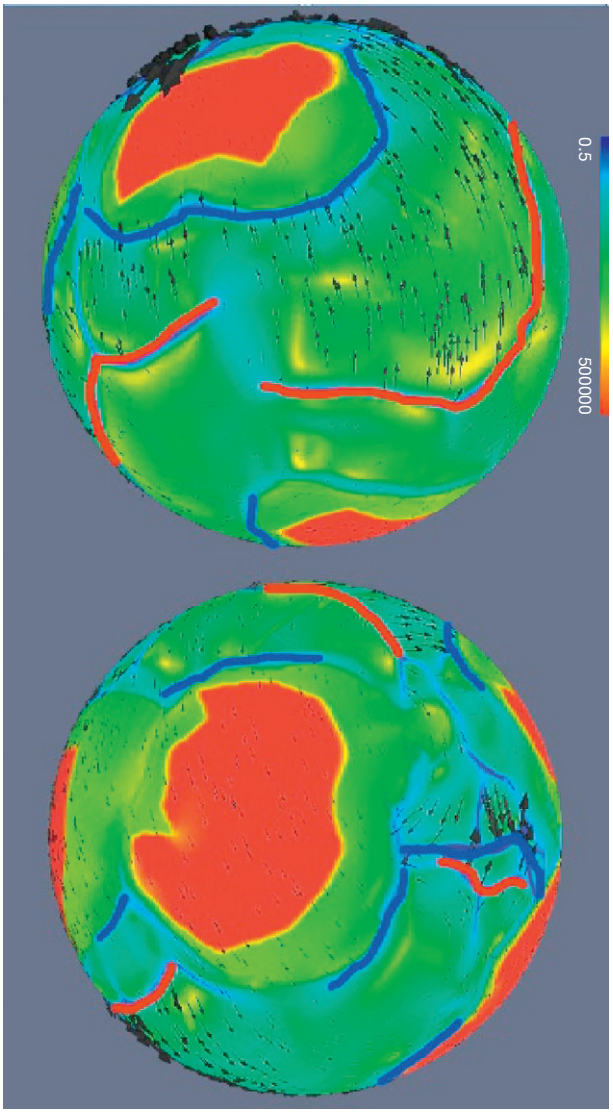


Figure 11 Viscosity and velocity fields at the surface of a 3-D spherical convection solution with platelike behavior and continental lithosphere. The red color is for the highest viscosity in the nucleus of continental lithosphere, surrounded by dark green weaker continental lithosphere. The light green is the viscosity of unstrained oceanic lithosphere, while blue represents low-viscosity regions where deformation localizes. The major divergent zones are highlighted by red curves, and the major convergent zones by blue curves. The surface velocity is close to that of rigid plates moving relative to each other. Courtesy of T. Rolf, N. Coltice and PJ. Tackley.

The topic is both controversial and somewhat academic. First, we do not have a decisive argument to explain why plate tectonics exists on Earth and not on other planets of the solar system (and particularly not on the Earth's sister planet, Venus). Second, the direct observations of the presence or absence of plate tectonics on super-Earths may also not happen in a close future. The debate on the propensity on plate tectonics on other planets illustrates our limits in the understanding of our own planet. The first papers on the subject had reached opposing conclusions about the likelihood of plate tectonics on super-Earths, inevitable for Valencia et al. (2007a) while less probable for O'Neill and Lenardic (2007).

The following studies, with all variables properly scaled, have concluded that plate tectonics is as favorable or more favorable on larger terrestrial planets than on Earth (Foley et al., 2012; Korenaga, 2010; Valencia and O'Connell, 2009; van Heck and Tackley, 2011). The stress level imposed by a more vigorous convection on larger planets (containing more radioactive elements and keeping its initial accretion heat a longer time) is higher. Combined with a thinner and weaker lithosphere, the internal stresses are more likely to induce the yielding of this lithosphere and to produce plate tectonic convection. However, other effects, namely, the surface temperature of the planet (Bercovici and Ricard, 2014; Foley et al., 2012) and the presence of water (Korenaga, 2010), have been shown equally or more important than the internal temperature for reaching plate tectonic convection. A hot temperature favors the healing of the lithosphere and makes plate tectonics less likely. Water lubricates the faults and makes plate tectonics more likely. This suggests that the climate of a planet controlling both the surface temperature and the water content may affect or even cause plate tectonics (Landuyt and Bercovici, 2009; Lenardic et al., 2008). The age of the planet since its accretion seems also an important parameter to consider (Lenardic and Crowley, 2012).

Acknowledgment

This chapter benefited from a very careful check of the equations by Frédéric Chambat and from the detailed and constructive reviews by Neil Ribe, Gerry Schubert, and Dave Bercovici. I also thank Ondrej Sramek, William Landuyt, Nicolas Coltice, Fabien Dubuffet, Thierry Alboussière, and others for their comments and suggestions.

References

Alboussiere T and Ricard Y (2013) Reflections on dissipation associated with thermal convection. *Journal of Fluid Mechanics* 725. http://dx.doi.org/10.1017/jfm.2013.241.

Alisic L, Gurnis M, Stadler G, Burstedde C, and Ghattas O (2012) Multi-scale dynamics and rheology of mantle flow with plates. *Journal of Geophysical Research* 117. http://dx.doi.org/10.1029/2012JB009234.

Anderson OL (1979) Evidence supporting the approximation $\gamma\rho = \text{const}$ for the Grüneisen parameter of the earth lower mantle. *Journal of Geophysical Research* 84: 3537–3542.

Anderson DL (2001) Geophysics – Top-down tectonics? *Science* 293: 2016–2018.

Austin NJ and Evans B (2007) Paleowattmeters: A scaling relation for dynamically recrystallized grain size. *Geology* 35(4): 343–346. http://dx.doi.org/10.1130/G23244A.1.

Auth C, Bercovici D, and Christensen UR (2003) Two-dimensional convection with a self-lubricating, simple-damage rheology. *Geophysical Journal International* 154: 783–800.

Badro J, Rueff JP, Vanko G, et al. (2004) Electronic transitions in perovskite: Possible nonconvecting layers in the lower mantle. *Science* 305: 383–386.

Bai WM, Vigny C, Ricard Y, and Froidevaux C (1992) On the origin of deviatoric stresses in the lithosphere. *Journal of Geophysical Research* 97: 11,729–11,737.

Balachandar S, Yuen DA, and Reuteler D (1992) Time-dependent 3-dimensional compressible convection with depth-dependent properties. *Geophysical Research Letters* 19: 2247–2250.

Balachandar S, Yuen DA, and Reuteler D (1993) Viscous and adiabatic heating effects in 3-dimensional compressible convection at infinite prandtl number. *Physics of Fluids A: Fluid Dynamics* 5: 2938–2945.

Barenblatt GI (1996) *Scaling, Self-similarity, and Intermediate Asymptotics*. Cambridge University Press.

Batchelor GK (1967) *An Introduction to Fluid Dynamics*. Cambridge University Press.

Bear J (1988) *Dynamics of Fluids in Porous Media*. New York: Dover publishers.

Beaumont C (1978) Evolution of sedimentary basins on a viscoelastic lithosphere – Theory and examples. *Geophysical Journal of the Royal Astronomical Society* 55: 471–497.

Becker TW and Boschi L (2002) A comparison of tomographic and geodynamic mantle models. *Geochemistry, Geophysics, Geosystems* 3: 1003. http://dx.doi.org/10.1029/2001GC000168.

Bercovici D (1993) A simple-model of plate generation from mantle flow. *Geophysical Journal International* 114: 635–650.

Bercovici D (1995) A source-sink model of the generation of plate-tectonics from non-Newtonian mantle flow. *Journal of Geophysical Research* 100: 2013–2030.

Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121.

Bercovici D and Karato S (2003) Whole-mantle convection and the transition-zone water filter. *Nature* 425: 39–44.

Bercovici D and Michaut C (2010) Two-phase dynamics of volcanic eruptions: Compaction, compression and the conditions for choking. *Geophysical Journal International* 182(2): 843–864. http://dx.doi.org/10.1111/j.1365-246X.2010.04674.x.

Bercovici D and Ricard Y (2003) Energetics of a two-phase model of lithospheric damage, shear localization and plate-boundary formation. *Geophysical Journal International* 152: 581–596.

Bercovici D and Ricard Y (2005) Tectonic plate generation and two-phase damage: Void growth versus grain size reduction. *Journal of Geophysical Research* 110.

Bercovici D and Ricard Y (2012) Mechanisms for the generation of plate tectonics by two-phase grain-damage and pinning. *Physics of the Earth and Planetary Interiors* 202–203: 27–55. http://dx.doi.org/10.1016/j.pepi.2012.05.003.

Bercovici D and Ricard Y (2013) Generation of plate tectonics with two-phase grain-damage and pinning: Source-sink model and toroidal flow. *Earth and Planetary Science Letters* 365: 275–288. http://dx.doi.org/10.1016/j.epsl.2013.02.002.

Bercovici D and Ricard Y (2014) Plate tectonics, damage and inheritance. *Nature* 508: 513–516.

Bercovici D, Ricard Y, and Richards MA (2000) The relation between mantle dynamics and plate tectonics: A primer. In: Richards MA, Gordon R, and VanderHilst R (eds.) *The History and Dynamics of Global Plate Motions*. *Geophysical Monograph Series*, vol. 21, pp. 5–46. Washington, DC: American Geophysical Union.

Bercovici D, Ricard Y, and Schubert G (2001a) A two-phase model of compaction and damage 1. General theory. *Journal of Geophysical Research* 106: 8887–8906.

Bercovici D, Ricard Y, and Schubert G (2001b) A two-phase model for compaction and damage 3. Applications to shear localization and plate boundary formation. *Journal of Geophysical Research* 106: 8925–8939.

Bercovici D, Schubert G, and Glatzmaier GA (1989a) 3-dimensional spherical models of convection in the earth's mantle. *Science* 244: 950–955.

Bercovici D, Schubert G, and Glatzmaier GA (1992) 3-dimensional convection of an infinite-Prandtl-number compressible fluid in a basally heated spherical-shell. *Journal of Fluid Mechanics* 239: 683–719.

Bercovici D, Schubert G, Glatzmaier GA, and Zebib A (1989b) 3-dimensional thermal-convection in a spherical-shell. *Journal of Fluid Mechanics* 206: 75–104.

Besse J and Courtillot V (1991) Revised and synthetic apparent polar wander paths of the African, Eurasian, North America and Indian plates, and true polar wander since 200 Ma. *Journal of Geophysical Research* 96: 4029–4050.

Birch F (1952) Elasticity and constitution of the Earth's interior. *Journal of Geophysical Research* 57: 227–286.

Bolton H and Masters G (2001) Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle. *Journal of Geophysical Research* 106: 13,527–13,540.

Booker JR (1976) Thermal-convection with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 76: 741–754.

Braun J, Chery J, Poliakov A, et al. (1999) A simple parameterization of strain localization in the ductile regime due to grain size reduction: A case study for olivine. *Journal of Geophysical Research* 104: 25,167–25,181.

Buffett BA, Garnero EJ, and Jeanloz R (2000) Sediments at the top of Earth's core. *Science* 290: 1338–1342.

Bullen KE (1940) The problem of the Earth's density variation. *Bulletin of the Seismological Society of America* 30: 235–250.

Bunge H, Hagelberg C, and Travis B (2003) Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography. *Geophysical Journal International* 152(2): 280–301. http://dx.doi.org/10.1046/j.1365-246X.2003.01823.x.

Bunge HP, Ricard Y, and Matas J (2001) Non-adiabaticity in mantle convection. *Geophysical Research Letters* 28: 879–882.

Bunge HP and Richards MA (1996) The origin of large scale structure in mantle convection: Effects of plate motions and viscosity stratification. *Geophysical Research Letters* 23: 2987–2990.

Bunge HP, Richards MA, and Baumgardner JR (1997) A sensitivity study of three-dimensional spherical mantle convection at 10(8) Rayleigh number: Effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *Journal of Geophysical Research* 102: 11991–12007.

Cadek O and Fleitout L (2003) Effect of lateral viscosity variations in the top 300 km on the geoid and dynamic topography. *Geophysical Journal International* 152: 566–580.

Carman P (1939) Permeability of saturated sands, soils and clays. *Journal of Agricultural Science* 29(Part 2): 262–273.

Chambat F, Benzoni-Gavage S, and Ricard Y (2014) Jump conditions and dynamic surface tension at permeable interfaces such as the inner core boundary. *Comptes Rendus Geoscience* 346: 110–118.

Chandrasekhar S (1969) *Ellipsoidal Figures of Equilibrium*. Haven: Yale Press.

Choblet G, Cadek O, Couturier F, and Dumoulin C (2007) OEDIPUS: A new tool to study the dynamics of planetary interiors. *Geophysical Journal International* 170(1): 9–30. http://dx.doi.org/10.1111/j.1365-246X.2007.03419.x.

Choblet G and Parmentier EM (2001) Mantle upwelling and melting beneath slow spreading centers: effects of variable rheology and melt productivity. *Earth and Planetary Science Letters* 184: 589–604.

Chopelas A and Boehler R (1992) Thermal expansivity in the lower mantle. *Geophysical Research Letters* 19: 1983–1986.

Christensen U (1984a) Convection with pressure-dependent and temperature-dependent non-Newtonian rheology. *Geophysical Journal of the Royal Astronomical Society* 77: 343–384.

Christensen UR (1984b) Heat-transport by variable viscosity convection and implications for the Earth's thermal evolution. *Physics of the Earth and Planetary Interiors* 35: 264–282.

Christensen UR (1996) The influence of trench migration on slab penetration into the lower mantle. *Earth and Planetary Science Letters* 140: 27–39.

Christensen UR (1997a) Some geodynamical effects of anisotropic viscosity. *Geophysical Journal of the Royal Astronomical Society* 91: 711–736.

Christensen UR (1997b) Influence of chemical buoyancy on the dynamics of slabs in the transition zone. *Journal of Geophysical Research* 102: 22,435–22,443.

Christensen UR and Harder H (1991) 3-d convection with variable viscosity. *Geophysical Journal International* 104: 213–226.

Christensen UR and Hofmann A (1994) Segregation of subducted oceanic crust in the convecting mantle. *Journal of Geophysical Research* 99: 19,867–19,884.

Christensen UR and Yuen DA (1984) The interaction of a subducting lithospheric slab with a chemical or phase-boundary. *Journal of Geophysical Research* 89: 4389–4402.

Colin P and Fleitout L (1990) Topography of the ocean-floor – Thermal evolution of the lithosphere and interaction of deep mantle heterogeneities with the lithosphere. *Geophysical Research Letters* 17: 1961–1964.

Coltice N, Moreira M, Hernlund J, and Labrosse S (2011) Crystallization of a basal magma ocean recorded by Helium and Neon. *Earth and Planetary Science Letters* 308(1–2): 193199. http://dx.doi.org/10.1016/j.epsl.2011.05.045.

Coltice N and Ricard Y (1999) Geochemical observations and one layer mantle convection. *Earth and Planetary Science Letters* 174: 125–137.

Coltice N, Rolf T, Tackley PJ, and Labrosse S (2012) Dynamic causes of the relation between area and age of the ocean floor. *Science* 336(6079): 335–338. http://dx.doi.org/10.1126/science.1219120.

Coltice N and Schmalzl J (2006) Mixing times in the mantle of the early Earth derived from 2-D and 3-D numerical simulations of convection. *Geophysical Research Letters* 33(23): L23304. http://dx.doi.org/10.1029/2006GL027707.

Connolly JAD and Podladchikov YY (1998) Compaction-driven fluid flow in viscoelastic rock. *Geodinamica Acta* 11: 55–84.

Conrad C and Gurnis M (2003) Seismic tomography, surface uplift, and the breakup of Gondwanaland: Integrating mantle convection backwards in time. *Geochemistry, Geophysics, Geosystems* 4. http://dx.doi.org/10.1029/2001GC000299.

Conrad CP and Husson L (2009) Influence of dynamic topography on sea level and its rate of change. *Lithosphere* 1(2): 110–120. http://dx.doi.org/10.1130/L32.1.

Corrieu V, Thoraval C, and Ricard Y (1995) Mantle dynamics and geoid Green functions. *Geophysical Journal International* 120: 516–523.

Crameri F, Tackley PJ, Meilick I, Gerya TV, and Kaus BJP (2012) A free plate surface and weak oceanic crust produce single-sided subduction on Earth. *Geophysical Research Letters* 39. http://dx.doi.org/10.1029/2011GL050046.

Cserepes L (1993) Effect of depth-dependent viscosity on the pattern of mantle convection. *Geophysical Research Letters* 20: 2091–2094.

Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press.

Davaille A (1999) Two-layer thermal convection in viscous fluids. *Journal of Fluid Mechanics* 379: 223–253.

Davaille A, Girard F, and LeBars M (2002) How to anchor hotspots in a convecting mantle? *Earth and Planetary Science Letters* 203: 621–634.

Davaille A and Jaupart C (1993) Transient high-rayleigh-number thermal-convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.

Davies GF (1980) Thermal histories of convective earth models and constraints on radiogenic heat-production in the earth. *Journal of Geophysical Research* 85: 2517–2530.

Davies GF (1999) *Dynamic Earth Plates, Plumes and Mantle Convection*, p 458. Cambridge, UK: Cambridge University Press.

Davies GF (2002) Stirring geochemistry in mantle convection models with stiff plates and slabs. *Geochimica et Cosmochimica Acta* 66: 3125–3142.

Davies DR, Goes S, Davies JH, et al. (2012) Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity. *Earth and Planetary Science Letters* 353: 253–269. http://dx.doi.org/10.1016/j.epsl.2012.08.016.

De Bresser J, Ter Heege J, and Spiers C (2001) Grain size reduction by dynamic recrystallization: can it result in major theological weakening? *International Journal of Earth Sciences* 90(1): 28–45.

de Groot SR and Mazur P (1984) *Non-Equilibrium Thermodynamics*. New York: Dover Publications.

Doin MP, Fleitout L, and Christensen U (1997) Mantle convection and stability of depleted and undepleted continental lithosphere. *Journal of Geophysical Research* 102: 2771–2787.

Dombre T, Frisch U, Greene JM, et al. (1986) Chaotic streamlines in the abc flows. *Journal of Fluid Mechanics* 167: 353–391.

Drew DA and Segel LA (1971) Averaged equations for 2-phase flows. *Studies in Applied Mathematics* 50: 205–220.

Drury MR and FitzGerald JD (1998) *The Earth's Mantle: Composition, Structure, and Evolution. Mantle Rheology: Insights from Laboratory Studies of Deformation and Phase Transition*. Cambridge, UK: Cambridge University Press.

Dubuffet F, Rabinowicz M, and Monnereau M (2000) Multiple scales in mantle convection. *Earth and Planetary Science Letters* 178: 351–366.

Dubuffet F, Yuen DA, and Rabinowicz M (1999) Effects of a realistic mantle thermal conductivity on the patterns of 3-D convection. *Earth and Planetary Science Letters* 171: 401–409.

Dziewonski AM and Anderson D (1981) Preliminary reference earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

England P and Mckenzie D (1982) A thin viscous sheet model for continental deformation. *Geophysical Journal of the Royal Astronomical Society* 70: 295–321.

Farnetani CG (1997) Excess temperature of mantle plumes: The role of chemical stratification across D″ *Geophysical Research Letters* 24: 1583–1586.

Ferrachat S and Ricard Y (1998) Regular vs, chaotic mantle mixing. *Earth and Planetary Science Letters* 155: 75–86.

Flament N, Gurnis M, and Müller RD (2013) A review of observations and models of dynamic topography. *Lithosphere* http://dx.doi.org/10.1130/L245.1.

Foley BJ and Becker TW (2009) Generation of plate-like behavior and mantle heterogeneity from a spherical, viscoplastic convection model. *Geochemistry, Geophysics, Geosystems* 10. http://dx.doi.org/10.1029/2009GC002378.

Foley BJ, Bercovici D, and Landuyt W (2012) The conditions for plate tectonics on super-Earths: Inferences from convection models with damage. *Earth and Planetary Science Letters* 331: 281–290. http://dx.doi.org/10.1016/j.epsl.2012.03.028.

Forte AM and Peltier WR (1987) Plate tectonics and aspherical earth structure – The importance of poloidal-toroidal coupling. *Journal of Geophysical Research* 92: 3645–3679.

Forte AM and Peltier WR (1991) Viscous flow models of global geophysical observables.1. forward problems. *Journal of Geophysical Research* 96: 20,131–20,159.

Fowler AC (1985) A mathematical model of magma transport in the asthenosphere. *Geophysical and Astrophysical Fluid Dynamics* 33: 63–96.

Fowler AC (1989) Generation and creep of magma in the earth. *SIAM Journal on Applied Mathematics* 49: 231–245.

Gable CW, Oconnell RJ, and Travis BJ (1991) Convection in 3 dimensions with surface plates – generation of toroidal flow. *Journal of Geophysical Research* 96: 8391–8405.

Garnero EJ and McNamara AK (2008) Structure and dynamics of Earth's lower mantle. *Science* 320(5876): 626–628. http://dx.doi.org/10.1126/science.1148028.

Gasperini P, Forno GD, and Boschi E (2004) Linear or non-linear rheology in the Earth's mantle: The prevalence of power-law creep in the postglacial isostatic readjustment of laurentia. *Geophysical Journal International* 157: 1297–1302.

Glatzmaier GA (1988) Numerical simulation of mantle convection: Time-dependent, three dimensional, compressible, spherical shell. *Geophysical and Astrophysical Fluid Dynamics* 43: 223–264.

Golabek GJ, Schmeling H, and Tackley PJ (2008) Earth's core formation aided by flow channelling instabilities induced by iron diapirs. *Earth and Planetary Science Letters* 271(1–4): 24–33. http://dx.doi.org/10.1016/j.epsl.2008.02.033.

Grigné C and Labrosse S (2001) Effects of continents on Earth cooling: Thermal blanketing and depletion in radioactive elements. *Geophysical Research Letters* 28: 2707–2710.

Grigne C, Labrosse S, and Tackley PJ (2005) Convective heat transfer as a function of wavelength: Implications for the cooling of the earth. *Journal of Geophysical Research, Solid Earth* 110.

Guillou L and Jaupart C (1995) On the effect of continents on mantle convection. *Journal of Geophysical Research* 100: 24,217–24,238.

Guilloufrottier L, Buttles J, and Olson P (1995) Laboratory experiments on the structure of subducted lithosphere. *Earth and Planetary Science Letters* 133: 19–34.

Gurnis M and Davies GF (1986) Numerical study of high Rayleigh number convection in a medium with depth-dependent viscosity. *Geophysical Journal of the Royal Astronomical Society* 85: 523–541.

Gurnis M and Hager BH (1988) Controls of the structure of subducted slabs. *Nature* 335: 317–321.

Gurnis M, Moresi L, and Mueller RD (2000) The history and dynamics of global plate motions. In: Richards MA, Gordon R, and vanderHilst R (eds.) *Models of mantle Convection Incorporating Plate Tectonics: The Australian Region since the Cretaceous*, pp. 211–238. Washington, DC: American Geophysical Union.

Haase R (1990) *Thermodynamics of Irreversible Processes*. New York: Dover Publications.

Hager BH and Clayton RW (1989) Constraints on the structure of mantle convection using seismic observation, flow models and the geoid. In: Peltier WR (ed.) *Mantle Convection, Plate Tectonics and Global Dynamics*. New York: Gordon and Breach Science Publishers.

Hager BH, Clayton RW, Richards MA, Comer RP, and Dziewonski AM (1985) Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* 313: 541–546.

Hager BH and O'Connell RJ (1981) A simple global model of plate tectonics and mantle convection. *Journal of Geophysical Research* 86: 4843–4867.

Hager BH and Oconnell RJ (1979) Kinematic models of large-scale flow in the Earth's mantle. *Journal of Geophysical Research* 84: 1031–1048.

Hansen U, Yuen DA, Kroening SE, and Larsen TB (1993) Dynamic consequences of depth-dependent thermal expansivity and viscosity on mantle circulations and thermal structure. *Physics of the Earth and Planetary Interiors* 77: 205–223.

Haskell NA (1937) The viscosity of the asthenosphere. *American Journal of Science* 33: 22–28.

Hernlund JW and Houser C (2008) The statistical distribution of seismic velocities in Earth's deep mantle. *Earth and Planetary Science Letters* 265(3–4): 423–437. http://dx.doi.org/10.1016/j.epsl.2007.10.042.

Hernlund JW and Jellinek AM (2010) Dynamics and structure of a stirred partially molten ultralow-velocity zone. *Earth and Planetary Science Letters* 296(1–2): 1–8. http://dx.doi.org/10.1016/j.epsl.2010.04.027.

Hernlund J, Thomas C, and Tackley P (2005) A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* 434(7035): 882–886. http://dx.doi.org/10.1038/nature03472.

Hewitt IJ (2010) Modelling melting rates in upwelling mantle. *Earth and Planetary Science Letters* 300(3–4): 264–274. http://dx.doi.org/10.1016/j.epsl.2010.10.010.

Hewitt IJ and Fowler AC (2008) Partial melting in an upwelling mantle column. *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 464(2097): 2467–2491. http://dx.doi.org/10.1098/rspa.2008.0045.

Hewitt IJ and Fowler AC (2009) Melt channelization in ascending mantle. *Journal of Geophysical Research* 114. http://dx.doi.org/10.1029/2008JB006185.

Hewitt JM, McKenzie DP, and Weiss NO (1975) Dissipative heating in convective flows. *Journal of Fluid Mechanics* 68: 721–738.

Hier-Majumder S, Ricard Y, and Bercovici D (2006) Role of grain boundaries in magma migration and storage. *Earth and Planetary Science Letters* 248: 735–749.

Hillert M (1965) On theory of normal and abnormal grain growth. *Acta Metallurgica* 13: 227–235.

Hiraga T, Tachibana C, Ohashi N, and Sano S (2010) Grain growth systematics for forsterite +/- enstatite aggregates: Effect of lithology on grain size in the upper mantle. *Earth and Planetary Science Letters* 291(1–4): 10–20. http://dx.doi.org/10.1016/j.epsl.2009.12.026.

Hirth G and Kolhstedt DL (1996) Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144: 93–108.

Hofmann A and Hart S (1978) Assessment of local and regional isotopic equilibrium in mantle. *Earth and Planetary Science Letters* 38(1): 44–62. http://dx.doi.org/10.1016/0012-821X(78)90125-5.

Hofmeister AM (1999) Mantle values of thermal conductivity and the geotherm from phonon lifetimes. *Science* 283: 1699–1706.

Holtzman B, Kohlstedt D, Zimmerman M, et al. (2003) Melt segregation and strain partitioning: Implications for seismic anisotropy and mantle flow. *Science* 301(5637): 1227–1230. http://dx.doi.org/10.1126/science.1087132.

Honda S, Yuen DA, Balachandar S, and Reuteler D (1993) 3-dimensional instabilities of mantle convection with multiple phase-transitions. *Science* 259: 1308–1311.

Husson L and Conrad CP (2006) Tectonic velocities, dynamic topography, and relative sea level. *Geophysical Research Letters* 33(18). http://dx.doi.org/10.1029/2006GL026834.

Irifune T and Ringwood AE (1987) Phase transformations in a harzburgite composition to 26 GPa: implications for dynamical behaviour of the subducting slab. *Earth and Planetary Science Letters* 86: 365–376.

Irifune T and Ringwood A (1993) Phase transformations in subducted oceanic crust and buoyancy relationships at depths of 600–800 km in the mantle. *Earth and Planetary Science Letters* 117: 101–110.

Ishii M and Tromp J (1999) Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle. *Science* 285(5431): 1231–1236. http://dx.doi.org/10.1126/science.285.5431.1231.

Ismail-Zadeh A, Schubert G, Tsepelev I, and Korotkii A (2004) Inverse problem of thermal convection: numerical approach and application to mantle plume restoration. *Physics of the Earth and Planetary Interiors* 145(1–4): 99–114. http://dx.doi.org/10.1016/j.pepi.2004.03.006.

Ito G, Shen Y, Hirth G, and Wolfe CJ (1999) Mantle flow, melting, and dehydration of the Iceland mantle plume. *Earth and Planetary Science Letters* 165: 81–96.

Jaroslow GE, Hirth G, and Dick HJB (1996) Abyssal peridotite mylonites: Implications for grain-size sensitive flow and strain localization in the oceanic lithosphere. *Tectonophysics* 256: 17–37.

Jarvis GT and Mckenzie DP (1980) Convection in a compressible fluid with infinite prandtl number. *Journal of Fluid Mechanics* 96: 515–583.

Jaupart C and Mareschal JC (1999) The thermal structure and thickness of continental roots. *Lithos* 48: 93–114.

Jeanloz R and Morris S (1987) Is the mantle geotherm subadiabatic. *Geophysical Research Letters* 14: 335–338.

Jellinek AM and Manga M (2002) The influence of a chemical boundary layer on the fixity, spacing and lifetime of mantle plumes. *Nature* 418: 760–763.

Joseph D (1966) Nonlinear stability of boussinesq equations by method of energy. *Archive for Rational Mechanics and Analysis* 22(3): 163–184.

Joseph DD and Renardy Y (1993) *Fundamentals of Two-Fluid Dynamics. Part 1: Mathematical Theory and Applications*. New York: Springer-Verlag.

Kameyama M and Ogawa M (2000) Transitions in thermal convection with strongly temperature-dependent viscosity in a wide box. *Earth and Planetary Science Letters* 180: 355–367.

Kameyama M, Yuen DA, and Fujimoto H (1997) The interaction of viscous heating with grain-size dependent rheology in the formation of localized slip zones. *Geophysical Research Letters* 24: 2523–2526.

Karato SI (1998) Seismic anisotropy in the deep mantle, boundary layers and the geometry of mantle convection. *Pure and Applied Geophysics* 151: 565–587.

Karato S (2008) *Deformation of Earth Materials: An Introduction to the Rheology of Solid Earth*. Cambridge, UK: Cambridge University Press.

Karato S, Bercovici D, Leahy G, Richard G, and Jing Z (2006) The transition-zone water filter model for global material circulation: Where do we stand? In: Jacobsen SD and van der Lee S (eds.) *Earth's Deep Water Cycle*. *Geophysical Monograph Series*, vol. 168, pp. 289–313. Washington, DC: AGU.

Karato SI, Paterson MS, and Gerald JDF (1986) Rheology of synthetic olivine aggregates – Influence of grain-size and water. *Journal of Geophysical Research* 91: 8151–8176.

Karato SI and Wu P (1993) Rheology of the upper mantle: A synthesis. *Science* 260: 771–778.

Katz RF, Spiegelman M, and Carbotte SM (2004) Ridge migration, asthenospheric flow and the origin of magmatic segmentation in the global mid-ocean ridge system. *Geophysical Research Letters* 31.

Kaus BJP, Muehlhaus H, and May DA (2010) A stabilization algorithm for geodynamic numerical simulations with a free surface. *Physics of the Earth and Planetary Interiors* 181(1–2): 12–20. http://dx.doi.org/10.1016/j.pepi.2010.04.007.

Kelemen PB, Hirth G, Shimizu N, Spiegelman M, and Dick HJB (1997) A review of melt migration processes in the adiabatically upwelling mantle beneath oceanic spreading ridges. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 355: 283–318.

Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283: 1881–1884.

Kellogg LH and Turcotte DL (1987) Homogenization of the mantle by convective mixing and diffusion. *Earth and Planetary Science Letters* 81: 371–378.

Kennett BLN (2001) *The Seismic Wavefield*. Cambridge, UK: Cambridge University Press.

Kennett B and Gorbatov A (2004) Seismic heterogeneity in the mantle – Strong shear wave signature of slabs from joint tomography. *Physics of the Earth and Planetary Interiors* 146(1–2): 87100. http://dx.doi.org/10.1016/j.pepi.2003.07.033, Superplume International Workshop, Tokyo Inst. Tech, Tokyo, Japan, Jan 28–31, 2002.

Kido M and Seno T (1994) Dynamic topography compared with residual depth anomalies in oceans and implications for age-depth curves. *Geophysical Research Letters* 21: 717–720.

Kincaid C and Olson P (1987) An experimental-study of subduction and slab migration. *Journal of Geophysical Research* 92: 13,832–13,840.

King SD, Gable CW, and Weinstein SA (1992) Models of convection-driven tectonic plates – A comparison of methods and results. *Geophysical Journal International* 109: 481–487.

King SD and Hager BH (1994) Subducted slabs and the geoid.1. Numerical experiments with temperature-dependent viscosity. *Journal of Geophysical Research* 99: 19,843–19,852.

Korenaga J (2003) Energetics of mantle convection and the fate of fossil heat. *Geophysical Research Letters* 30.

Korenaga J (2010) On the likelihood of plate tectonics on super Earth's: does size matter? *Astrophysical Journal Letters* 725(1): L43–L46. http://dx.doi.org/10.1088/2041-8205/725/1/L43.

Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* 450(7171): 866–869. http://dx.doi.org/10.1038/nature06355.

Lago B and Rabinowicz M (1984) Admittance for convection in a layered spherical shell. *Geophysical Journal of the Royal Astronomical Society* 77(2): 461–482.

Lambeck K and Johnston P (1998) *The Earth's mantle: composition, structure, and evolution*. *The viscosity of the mantle*. Cambridge, UK: Cambridge University Press, pp. 461–502.

Landau L and Lifchitz E (1980) *An Introduction to Fluid Dynamics*. Oxford: Pergamon Press.

Landau L and Lifchitz E (2000) *Theory of elasticity*. London: Butterworth-Heinemann.

Landuyt W and Bercovici D (2009) Variations in planetary convection via the effect of climate on damage. *Earth and Planetary Science Letters* 277(1–2): 29–37. http://dx.doi.org/10.1016/j.epsl.2008.09.034.

Larsen TB, Malevsky AV, Yuen DA, and Smedsmo JL (1993) Temperature-dependent Newtonian and non-Newtonian convection – Implications for lithospheric processes. *Geophysical Research Letters* 20: 2595–2598.

Larsen TB and Yuen DA (1997) Fast plumeheads: Temperature-dependent versus non-Newtonian rheology. *Geophysical Research Letters* 24: 1995–1998.

Lasaga A (1998) *Kinetics Theory in the Earth Sciences*. Princeton, NJ: Princeton University Press.

LeBars M and Davaille A (2002) Stability of thermal convection in two superimposed miscible viscous fluids. *Journal of Fluid Mechanics* 471: 339–363.

Lee T, Papanastassiou DA, and Wasserburg GL (1976) Demonstration of ^{26} mg excess in allende and evidence for al^{26}. *Geophysical Research Letters* 3: 41–44.

Lemery C, Ricard Y, and Sommeria J (2000) A model for the emergence of thermal plumes in Rayleigh–Bénard convection at infinite Prandtl number. *Journal of Fluid Mechanics* 414: 225–250.

Lenardic A (1998) On the partitioning of mantle heat loss below oceans and continents over time and its relationship to the Archaean paradox. *Geophysical Journal International* 134: 706–720.

Lenardic A and Crowley JW (2012) On the notion of well-defined tectonic regimes for terrestrial planets in this solar system and others. *Astrophysical Journal* 755(2): 132. http://dx.doi.org/10.1088/0004-637X/755/2/132.

Lenardic A, Jellinek AM, and Moresi LN (2008) A climate change induced transition in the tectonic style of a terrestrial planet. *Earth and Planetary Science Letters* 271: 33–42.

Lestunff Y and Ricard Y (1995) Topography and geoid due to lithospheric mass anomalies. *Geophysical Journal International* 122: 982–990.

Lister JR and Kerr RC (1991) Fluid-mechanical models of crack-propagation and their application to magma transport in dykes. *Journal of Geophysical Research, Solid Earth* 96: 10,049–10,077.

Lithgow-Bertelloni C and Gurnis M (1997) Cenozoic subsidence and uplift of continents from time-varying dynamic topography. *Geology* 25: 735–738.

Lithgow-Bertelloni C and Guynn JH (2004) Origin of the lithospheric stress field. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2003JB002467.

Lithgow-Bertelloni C and Richards MA (1998) Dynamics of cenozoic and mesozoic plate motions. *Reviews of Geophysics* 36: 27–78.

Lithgow-Bertelloni C and Silver PG (1998) Dynamic topography, plate driving forces and the African superswell. *Nature* 395: 269–272.

Liu L and Gurnis M (2008) Simultaneous inversion of mantle properties and initial conditions using an adjoint of mantle convection. *Journal of Geophysical Research* 113(B8). http://dx.doi.org/10.1029/2008JB005594.

Lyubetskaya T and Korenaga J (2007) Chemical composition of Earth's primitive mantle and its variance, 1, method and results. *Journal of Geophysical Research* 112(B3): B03212.

Machetel P and Weber P (1991) Intermittent layered convection in a model mantle with an endothermic phase change at 670 km. *Nature* 350: 55–57.

Malevsky AV and Yuen DA (1992) Strongly chaotic non-Newtonian mantle convection. *Geophysical and Astrophysical Fluid Dynamics* 65: 149–171.

Malvern L (1969) *Introduction to the Mechanics of a Continuum Medium*. *Prentice-Hall Series in Engineering of the Physical Sciences*. Englewood, NJ: Prentice-Hall.

Massol H, Jaupart C, and Pepper DW (2001) Ascent and decompression of viscous vesicular magma in a volcanic conduit. *Journal of Geophysical Research* 106: 16,223–16,240.

Masters G, Laske G, Bolton H, and Dziewonski A (2000) The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure. In: Karato S, Forte AM, Liebermann RC, Masters G, and Stixrude L (eds.) *Earth's Deep Interior*. *Geophysical Monograph Series*, vol. 117. Washington, DC: AGU.

McDonough WF and Sun SS (1995) The composition of the earth. *Chemical Geology* 120: 223–253.

McKenzie D (1984) The generation and compaction of partially molten rock. *Journal of Petrology* 25: 713–765.

McNamara AK, Karato SI, and van Keken PE (2001) Localisation of dislocation creep in the lower mantle: Implications for the origin of seismic anisotropy. *Earth and Planetary Science Letters* 191: 85–99.

Michaut C and Bercovici D (2009) A model for the spreading and compaction of two-phase viscous gravity currents. *Journal of Fluid Mechanics* 630: 299–329. http://dx.doi.org/10.1017/S0022112009006612.

Michaut C, Bercovici D, and Sparks RSJ (2009) Ascent and compaction of gas rich magma and the effects of hysteretic permeability. *Earth and Planetary Science Letters* 282(1–4): 258–267. http://dx.doi.org/10.1016/j.epsl.2009.03.026.

Michaut C, Ricard Y, Bercovici D, and Sparks RSJ (2013) Eruption cyclicity at silicic volcanoes potentially caused by magmatic gas waves. *Nature Geoscience* 6(10): 856–860. http://dx.doi.org/10.1038/NGEO1928.

Mitrovica JX and Forte AM (1997) Radial profile of mantle viscosity: Results from the joint inversion of convection and postglacial rebound observables. *Journal of Geophysical Research* 102: 2751–2769.

Monnereau M and Quéré S (2001) Spherical shell models of mantle convection with tectonic plates. *Earth and Planetary Science Letters* 184: 575–588.

Moresi L and Solomatov V (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic styles of the Earth and Venus. *Geophysical Journal International* 133: 669–682.

Moucha R, Forte AM, Mitrovica JX, et al. (2008) Dynamic topography and long-term sea-level variations: There is no such thing as a stable continental platform. *Earth and Planetary Science Letters* 271(1–4): 101108. http://dx.doi.org/10.1016/j.epsl.2008.03.056.

Muhlhaus HB, Moresi L, and Cada M (2004) Emergent anisotropy and flow alignment in viscous rocks. *Pure and Applied Geophysics* 161: 2451–2463.

Munk WH and MacDonald GJF (1960) *The Rotation of the Earth: A Geophysical Discussion*. New York: Cambridge University Press.

Murakami T, Hirose K, Kawamura K, Sata N, and Ohishi Y (2004) Post-perovskite phase transition in $MgSiO_3$. *Science* 304: 855–858.

Murnaghan FD (1951) *Finite Deformation of an Elastic Solid*. New York: Wiley.

Nakagawa T and Tackley PJ (2006) Three-dimensional structures and dynamics in the deep mantle: Effects of post-perovskite phase change and deep mantle layering. *Geophysical Research Letters* 33: L12S11. http://dx.doi.org/1029/2006GL025719.

Nataf HC and Richter FM (1982) Convection experiments in fluids with highly temperature-dependent viscosity and the thermal evolution of the planets. *Physics of the Earth and Planetary Interiors* 29: 320–329.

Ni S, Tan E, Gurnis M, and Helmberger D (2002) Sharp sides to the African superplume. *Science* 296(5574): 1850–1852. http://dx.doi.org/10.1126/science.1070698.

Nield DA (1964) Surface tension and buoyancy effects in cellular convection. *Journal of Fluid Mechanics* 19: 341–352.

Nye JF (1953) The flow law of ice from measurements in glacier tunnels, laboratory experiments and the Jungfraufirn borehole experiment. *Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 219: 477–489.

O'Neill C and Lenardic A (2007) Geological consequences of super-sized Earths. *Geophysical Research Letters* 34(19). http://dx.doi.org/10.1029/2007GL030598.

Ogawa M, Schubert G, and Zebib A (1991) Numerical simulations of 3-dimensional thermal-convection in a fluid with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 233: 299–328.

Ogura Y and Phillips N (1962) Scale analysis of deep and shallow convection in the atmosphere. *Journal of Atmospheric Science* 19(2): 173–179. http://dx.doi.org/10.1175/1520-0469.

Olson P, Yuen DA, and Balsiger D (1984) Mixing of passive heterogeneities by mantle convection. *Journal of Geophysical Research* 89: 425–436.

Ottino JM (1989) *The Kinematics of Mixing: Stretching, Chaos and Transport*. New York: Cambridge University Press.

Panasyuk S, Hager B, and Forte A (1996) Understanding the effects of mantle compressibility on geoid kernels. *Geophysical Journal International* 124: 121–133.

Parmentier EM, Sotin C, and Travis BJ (1994) Turbulent 3-d thermal-convection in an infinite Prandtl number, volumetrically heated fluid – Implications for mantle dynamics. *Geophysical Journal International* 116: 241–251.

Parmentier EM, Turcotte DL, and Torrance KE (1976) Studies of finite-amplitude non-Newtonian thermal convection with application to convection in Earth's mantle. *Journal of Geophysical Research* 81: 1839–1846.

Parsons B and Daly S (1983) The relationship between surface topography, gravity anomaly and the temperature structure of convection. *Journal of Geophysical Research* 88: 1129–1144.

Peltier WR (1989) Mantle convection: Plate tectonics and Global geodynamics. In: *Mantle Viscosity*, pp. 389–478. New York: Gordon and Breach Science Publishers.

Peltier WR (1996) Mantle viscosity and ice-age ice sheet topography. *Science* 273: 1359–1364.

Peltier WR and Solheim LP (1992) Mantle phase transition and layered chaotic convection. *Geophysical Research Letters* 19: 321–324.

Poirier JP (1991) *Introduction to the Physics of the Earth's Interior*. Cambridge: Cambridge University Press.

Poirier JP and Tarantola A (1998) A logarithmic equation of state. *Physics of the Earth and Planetary Interiors* 109: 1–8.

Pouilloux L, Kaminski E, and Labrosse S (2007) Anisotropic rheology of a cubic medium and implications for geological materials. *Geophysical Journal International* 170(2): 876–885. http://dx.doi.org/10.1111/j.1365-246X.2007.03461.x.

Rabinowicz M, Ricard Y, and Gregoire M (2002) Compaction in a mantle with a very small melt concentration: Implications for the generation of carbonatitic and carbonate-bearing high alkaline mafic melt impregnations. *Earth and Planetary Science Letters* 203: 205–220.
Ranalli G (1995) *Rheology of the Earth*. Dordrecht: Kluwer academic publishers.
Ribe NM (1985a) The deformation and compaction of partial molten zones. *Geophysical Journal of the Royal Astronomical Society* 83: 487–501.
Ribe NM (1985b) The generation and composition of partial melts in the Earth's mantle. *Earth and Planetary Science Letters* 73: 361–376.
Ribe NM, Stutzmann E, Ren Y, and van der Hilst R (2007) Buckling instabilities of subducted lithosphere beneath the transition zone. *Earth and Planetary Science Letters* 254(1–2): 173–179. http://dx.doi.org/10.1016/j.epsl.2006.11.028.
Ricard Y and Bercovici D (2003) Two-phase damage theory and crustal rock failure: The theoretical 'void' limit, and the prediction of experimental data. *Geophysical Journal International* 155: 1057–1064.
Ricard Y and Bercovici D (2009) A continuum theory of grain size evolution and damage. *Journal of Geophysical Research* 114(B01204). http://dx.doi.org/10.1029/2007JB005491.
Ricard Y, Bercovici D, and Schubert G (2001) A two-phase model for compaction and damage 2. Applications to compaction, deformation, and the role of interfacial surface tension. *Journal of Geophysical Research* 106: 8907–8924.
Ricard Y and Coltice N (2004) Geophysical and geochemical models of mantle convection: Successes and future challenges. In: *The state of the planet: Frontiers and challenges in geophysics*. *Geophysical Monograph Series*, vol. 150, pp. 59–68. Washington, DC: AGU.
Ricard Y, Doglioni C, and Sabadini R (1991) Differential rotation between lithosphere and mantle – A consequence of lateral mantle viscosity variations. *Journal of Geophysical Research* 96: 8407–8415.
Ricard Y, Fleitout L, and Froidevaux C (1984) Geoid heights and lithospheric stresses for a dynamic Earth. *Annals of Geophysics* 2: 267–286.
Ricard Y, Mattern E, and Matas J (2005) Earth's deep mantle: Structure, composition, and evolution. In: van der Hilst RD, Bass JD, Matas J, and Trampert J (eds.) *Synthetic Tomographic Images of Slabs from Mineral Physics*. *Geophysical Monograph Series*, pp. 285–302. Washington, DC: AGU.
Ricard Y, Richards M, Lithgow-Bertelloni C, and Stunff YL (1993a) A geodynamic model of mantle density heterogeneity. *Journal of Geophysical Research* 98: 21,895–21,909.
Ricard Y, Spada G, and Sabadini R (1993b) Polar wander of a dynamic earth. *Geophysical Journal International* 113: 284–298.
Ricard Y, Sramek O, and Dubuffet F (2009) A multi-phase model of runaway core–mantle segregation in planetary embryos. *Earth and Planetary Science Letters* 284(1–2): 144–150. http://dx.doi.org/10.1016/j.epsl.2009.04.021.
Ricard Y and Vigny C (1989) Mantle dynamics with induced plate-tectonics. *Journal of Geophysical Research* 94: 17,543–17,559.
Ricard Y, Vigny C, and Froidevaux C (1989) Mantle heterogeneities, geoid and plate motion: a monte carlo inversion. *Journal of Geophysical Research* 94: 13,739–13,754.
Richards M (1991) Hotspots and the case for a high viscosity lower mantle. In: Sabadini R, et al. (eds.) *Glacial Isostasy, Sea-Level and Mantle Rheology*, pp. 571–587. Dordrecht: Kluwer Academic Publishers.
Richards MA and Engebretson DC (1992) Large-scale mantle convection and the history of subduction. *Nature* 355: 437–440.
Richards MA and Hager BH (1984) Geoid anomalies in a dynamic earth. *Journal of Geophysical Research* 89: 5987–6002.
Richards MA, Ricard Y, Lithgow-Bertelloni G Spada, and Sabadini R (1997) An explanation for Earth's long-term rotational stability. *Science* 275: 372–375.
Richards MA, Yang WS, Baumgardner JR, and Bunge HP (2001) Role of a low-viscosity zone in stabilizing plate tectonics: Implications for comparative terrestrial planetology. *Geochemistry, Geophysics, Geosystems* 2.
Richter FM (1978) Experiments on the stability of convection rolls in fluids whose viscosity depends on temperature. *Journal of Fluid Mechanics* 89: 553–560.
Richter FM and Mckenzie D (1984) Dynamical models for melt segregation from a deformable matrix. *Journal of Geology* 92: 729–740.
Ricolleau A, Fiquet G, Perillat J, Daniel I, Menguy N, Cardon H, Addad A, Vanni C, and Guignot N (2004) The fate of subducted basaltic crust in the earth's lower mantle: an experimental petrological study. *Eos, Transactions of the American Geophysical Union* 85(47), Fall Meet. Suppl., Abstract U33B-02.
Riley GN, Kohlstedt DL, and Richter FM (1990) Melt migration in a silicate liquid-olivine system – An experimental test of compaction theory. *Geophysical Research Letters* 17: 2101–2104.
Ringwood AE and Irifune T (1988) Nature of the 650-km seismic discontinuity – Implications for mantle dynamics and differentiation. *Nature* 331: 131–136.
Roberts PH (1967) Convection in horizontal layers with internal heat generation: Theory. *Journal of Fluid Mechanics* 30: 33–49.
Rolf T, Coltice N, and Tackley PJ (2012) Linking continental drift, plate tectonics and the thermal state of the Earth's mantle. *Earth and Planetary Science Letters* 351: 134–146. http://dx.doi.org/10.1016/j.epsl.2012.07.011.
Rolf T and Tackley PJ (2011) Focussing of stress by continents in 3D spherical mantle convection with self-consistent plate tectonics. *Geophysical Research Letters* 38. http://dx.doi.org/10.1029/2011GL048677.
Ross M and Schubert G (1986) Tidal dissipation in a viscoelastic planet. *Journal of Geophysical Research* 91(B4): D447–D452. http://dx.doi.org/10.1029/JB091iB04p0D447.
Rozel A, Ricard Y, and Bercovici D (2011) A thermodynamically self-consistent damage equation for grain size evolution during dynamic recrystallization. *Geophysical Journal International* 184(2): 719–728. http://dx.doi.org/10.1111/j.1365-246X.2010.04875.x.
Rudge JF, Bercovici D, and Spiegelman M (2011) Disequilibrium melting of a two phase multicomponent mantle. *Geophysical Journal International* 184(2): 699–718. http://dx.doi.org/10.1111/j.1365-246X.2010.04870.x.
Rudnick RL and Fountain DM (1995) Nature and composition of the continental crust – A lower crustal perspective. *Reviews of Geophysics* 33: 267–309.
Runcorn SK (1964) Satellite gravity measurements and the laminar viscous flow model of the Earth's mantle. *Journal of Geophysical Research* 69: 4389.
Sabadini R and Yuen DA (1989) Mantle stratification and long-term polar wander. *Nature* 339: 373–375.
Saltzer RL, Stutzmann E, and van der Hilst RD (2004) Poisson's ratio in the lower mantle beneath Alaska: Evidence for compositional heterogeneity. *Journal of Geophysical Research* 109.
Samuel H and Farnetani CG (2003) Thermochemical convection and helium concentrations in mantle plumes. *Earth and Planetary Science Letters* 207: 39–56.
Schmalzl J, Houseman GA, and Hansen U (1996) Mixing in vigorous, time-dependent three-dimensional convection and application to Earth's mantle. *Journal of Geophysical Research* 101: 21,847–21,858.
Schmeling H (2000) *Physics and Chemistry of Partially Molten Rocks*. *Partial Melting and Melt Migration in a Convecting Mantle*. Dordrecht: Kluwer Academic Publisher.
Schmeling H and Jacoby WR (1981) On modeling the lithosphere in mantle convection with non-linear rheology. *Journal of Geophysics* 50: 89–100.
Schmeling H, Marquart G, and Ruedas T (2003) Pressure- and temperature-dependent thermal expansivity and the effect on mantle convection and surface observables. *Geophysical Journal International* 154: 224–229.
Schmeling H, et al. (2008) A benchmark comparison of spontaneous subduction models – Towards a free surface. *Physics of the Earth and Planetary Interiors* 171(1–4): 198–223. http://dx.doi.org/10.1016/j.pepi.2008.06.028.
Schubert G, Stevenson D, and Cassen P (1980) Whole planet cooling and the radiogenic heat-source contents of the earth and moon. *Journal of Geophysical Research* 85: 2531–2538.
Schubert G, Turcotte D, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge, UK: Cambridge University Press.
Schubert G, Yuen DA, and Turcotte DL (1975) Role of phase-transitions in a dynamic mantle. *Geophysical Journal of the Royal Astronomical Society* 42: 705–735.
Scott DR and Stevenson DJ (1984) Magma solitons. *Geophysical Research Letters* 11: 9283–9296.
Scott DR and Stevenson DJ (1986) Magma ascent by porous flow. *Journal of Geophysical Research* 91: 9283–9296.
Scott DR and Stevenson DJ (1989) A self-consistent model of melting, magma migration and buoyancy-driven circulation beneath mid-ocean ridges. *Journal of Geophysical Research* 94: 2973–2988.
Shephard GE, Mueller RD, Liu L, and Gurnis M (2010) Miocene drainage reversal of the Amazon river driven by plate-mantle interaction. *Nature Geoscience* 3(12): 870–875. http://dx.doi.org/10.1038/NGEO1017.
Simpson G, Spiegelman M, and Weinstein MI (2010a) A multiscale model of partial melts: 1. Effective equations. *Journal of Geophysical Research* 115. http://dx.doi.org/10.1029/2009JB006375.
Simpson G, Spiegelman M, and Weinstein MI (2010b) A multiscale model of partial melts: 2. Numerical results. *Journal of Geophysical Research* 115. http://dx.doi.org/10.1029/2009JB006376.
Sleep NH (1988) Tapping of melt by veins and dikes. *Journal of Geophysical Research* 93(B9): 10255–10272. http://dx.doi.org/10.1029/JB093iB09p10255.
Solomatov VS (1995) Scaling of temperature-dependent and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.
Spada G, Ricard Y, and Sabadini R (1992a) Excitation of true polar wander by subduction. *Nature* 360: 452–454.

Spada G, Sabadini R, Yuen DA, and Ricard Y (1992b) Effects on postglacial rebound from the hard rheology in the transition zone. *Geophysical Journal International* 109: 683–700.

Spear FS (1993) *Metamorphic Phase Equilibria and Pressure–Temperature–Time Paths*, p. 799. Washington, DC: Mineralogical Society of America.

Spiegel E and Veronis G (1960) On the Boussinesq approximation for a compressible fluid. *Astrophysical Journal* 131(2): 442–447. http://dx.doi.org/10.1086/146849.

Spiegelman M and Kelemen PB (2003) Extreme chemical variability as a consequence of channelized melt transport. *Geochemistry, Geophysics, Geosystems* 4.

Spiegelman M, Kelemen PB, and Aharonov E (2001) Causes and consequences of flow organization during melt transport: The reaction infiltration instability in compactible media. *Journal of Geophysical Research* 106: 2061–2077.

Spiegelman M and McKenzie D (1987) Simple 2-D models for melt extraction at midocean ridges and island arcs. *Earth and Planetary Science Letters* 83(1–4): 137–152. http://dx.doi.org/10.1016/0012-821X(87)90057-4.

Sramek O, Milelli L, Ricard Y, and Labrosse S (2012) Thermal evolution and differentiation of planetesimals and planetary embryos. *Icarus* 217(1): 339–354. http://dx.doi.org/10.1016/j.icarus.2011.11.021.

Sramek O, Ricard Y, and Bercovici D (2007) Simultaneous melting and compaction in deformable two phase media. *Geophysical Journal International* 168(3): 964–982.

Sramek O, Ricard Y, and Dubuffet F (2010) A multiphase model of core formation. *Geophysical Journal International* 181(1): 198–220. http://dx.doi.org/10.1111/j.1365-246X.2010.04528.x.

Stacey FD (1977) Application of thermodynamics to fundamental earth physics. *Geophysical Surveys* 3: 175–204.

Stacey FD and Davis PM (2004) High pressure equations of state with applications to the lower mantle and core. *Physics of the Earth and Planetary Interiors* 142: 137–184.

Stadler G, Gurnis M, Burstedde C, et al. (2010) The dynamics of plate tectonics and mantle flow: From local to global scales. *Science* 329(5995): 1033–1038. http://dx.doi.org/10.1126/science.1191223.

Stein C, Schmalzl J, and Hansen U (2004) The effect of rheological parameters on plate behaviour in a self-consistent model of mantle convection. *Physics of the Earth and Planetary Interiors* 142: 225–255.

Steinberger B and O'Connell RJ (1998) Advection of plumes in mantle flow: Implications for hotspot motion, mantle viscosity and plume distribution. *Geophysical Journal International* 132: 412–434.

Stevenson DJ (1989) Spontaneous small-scale melt segregation in partial melts undergoing deformation. *Geophysical Research Letters* 16: 1067–1070.

Stixrude L and Lithgow-Bertelloni C (2005) Thermodynamics of mantle minerals I. Physical properties. *Geophysical Journal International* 162(2): 610–632. http://dx.doi.org/10.1111/j.1365-246X.2005.02642.x.

Tackley P (1995) On the penetration of an endothermic phase transition by upwellings and downwellings. *Journal of Geophysical Research* 100: 15,477–15,488.

Tackley PJ (1996) Effects of strongly variable viscosity on three-dimensional compressible convection in planetary mantles. *Journal of Geophysical Research* 101: 3311–3332.

Tackley PJ (1998) Self-consistent generation of tectonic plates in three-dimensional mantle convection. *Earth and Planetary Science Letters* 157: 9–22.

Tackley P (2000a) The quest for self-consistent incorporation of plate tectonics in mantle convection. In: Richards MA, Gordon R, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*. *AGU Geophysical Monograph Series*, vol. 121, pp. 47–72. Washington, DC: AGU.

Tackley PJ (2000b) Mantle convection and plate tectonics: Toward an integrated physical and chemical theory. *Science* 288: 2002–2007.

Tackley PJ (2000c) Self consistent generation of tectonic plates in time-dependent, three dimensional mantle convection simulations, part 1: Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* 1.

Tackley PJ (2000d) Self consistent generation of tectonic plates in time-dependent, three dimensional mantle convection simulations, part 2: Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems* 1.

Tackley PJ (2008) Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid. *Physics of the Earth and Planetary Interiors* 171(1–4): 7–18. http://dx.doi.org/10.1016/j.pepi.2008.08.005 (special issue).

Tackley PJ, Ammann M, Brodholt JP, Dobson DP, and Valencia D (2013) Mantle dynamics in super-Earth's: Post-perovskite rheology and self-regulation of viscosity. *Icarus* 225(1): 50–61. http://dx.doi.org/10.1016/j.icarus.2013.03.013.

Tackley PJ, Stevenson DJ, Glatzmaier GA, and Schubert G (1993) Effects of an endothermic phase-transition at 670 km depth in a spherical model of convection in the Earth's mantle. *Nature* 361: 699–704.

Tackley PJ and Xie SX (2002) The thermochemical structure and evolution of Earth's mantle: Constraints and numerical models. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 360: 2593–2609.

Tao WC and Oconnell RJ (1993) Deformation of a weak subducted slab and variation of seismicity with depth. *Nature* 361: 626–628.

Tauzin B, Debayle E, and Wittlinger G (2010) Seismic evidence for a global low-velocity layer within the Earth's upper mantle. *Nature Geoscience* 3(10): 718–721. http://dx.doi.org/10.1038/NGEO969.

Toth J and Gurnis M (1998) Dynamics of subduction initiation at preexisting fault zones. *Journal of Geophysical Research* 103: 18,053–18,067.

Toussaint V, Carriere P, Scott J, and Gence JN (2000) Spectral decay of a passive scalar in chaotic mixing. *Physics of Fluids* 12: 2834–2844.

Tozer DC (1972) The present thermal state of the terrestrial planets. *Physics of the Earth and Planetary Interiors* 6: 182–197.

Tritton DJ (1988) *Physical Fluid Dynamics*. Oxford: Oxford University Press.

Trompert R and Hansen U (1998a) Mantle convection simulations with rheologies that generate plate-like behaviour. *Nature* 395: 686–689.

Trompert RA and Hansen U (1998b) On the Rayleigh number dependence of convection with a strongly temperature-dependent viscosity. *Physics of Fluids* 10: 351–360.

Trubienko O, Fleitout L, Garaud J-D, and Vigny C (2013) Interpretation of interseismic deformations and the seismic cycle associated with large subduction earthquakes. *Tectonophysics* 589: 126–141. http://dx.doi.org/10.1016/j.tecto.2012.12.027.

Trull T and Kurtz M (1993) Experimental measurements of the He-3 and He-4 mobility in olivine and clinopyroxene at magmatic temperatures. *Geochimica et Cosmochimica Acta* 57(6): 1313–1324. http://dx.doi.org/10.1016/0016-7037(93)90068-8.

Turcotte DL and Morgan JP (1992) Mantle flow and melt generation at mid-ocean ridges. In: Phipps Morgan J, Blackmann DK, and Simpson JM (eds.) *The Physics of Magma Migration and Mantle Flow Beneath a Mid-Ocean Ridge*, pp. 155–182. New York: AGU.

Turcotte DL and Schubert G (1982) *Geodynamics: Applications of Continuum Physics to Geological Problems*, pp. 1–450. New York: Wiley.

Urey HC (1951) The origin and development of the earth and other terrestrial planets. *Geochimica et Cosmochimica Acta* 1: 209–277.

Valencia D and O'Connell RJ (2009) Convection scaling and subduction on Earth and super-Earth's. *Earth and Planetary Science Letters* 286(3–4): 492–502. http://dx.doi.org/10.1016/j.epsl.2009.07.015.

Valencia D, O'Connell RJ, and Sasselov DD (2007a) Inevitability of plate tectonics on super-Earths. *Astrophysical Journal* 670(1, Part 2): L45–L48. http://dx.doi.org/10.1086/524012.

Valencia D, Sasselov DD, and O'Connell RJ (2007b) Radius and structure models of the first super-earth planet. *Astrophysical Journal* 656(1, Part 1): 545–551. http://dx.doi.org/10.1086/509800.

van der Hilst RD, Widiyantoro S, and Engdahl ER (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386: 578–584.

van Heck HJ and Tackley PJ (2008) Planforms of self-consistently generated plates in 3D spherical geometry. *Geophysical Research Letters* 35(19). http://dx.doi.org/10.1029/2008GL035190.

van Heck HJ and Tackley PJ (2011) Plate tectonics on super-Earths: Equally or more likely than on Earth. *Earth and Planetary Science Letters* 310(3–4): 252–261. http://dx.doi.org/10.1016/j.epsl.2011.07.029.

van Keken PE, Hauri EH, and Ballentine CJ (2002) Mantle mixing: The generation, preservation, and destruction of chemical heterogeneity. *Annual Review of Earth and Planetary Sciences* 30: 493–525.

van Keken P, Yuen DA, and Vandenberg A (1992) Pulsating diapiric flows – Consequences of vertical variations in mantle creep laws. *Earth and Planetary Science Letters* 112: 179–194.

Vanderwal D, Chopra P, Drury M, and Fitz Gerald J (1993) Relationships between dynamically recrystallized grain-size and deformation conditions in experimentally deformed olivine rocks. *Geophysical Research Letters* 20(14): 1479–1482. http://dx.doi.org/10.1029/93GL01382.

Vanyo J, Wilde P, Cardin P, and Olson P (1995) Experiments on precessing flows in the earth's liquid core. *Geophysical Journal International* 121: 136–142.

Vinet P, Ferrante J, Rose JH, and Smith JR (1987) Compressibility of solids. *Journal of Geophysical Research* 92: 9319–9325.

Walker J, Hays P, and Kasting J (1981) A negative feedback mechanism for long-term stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86(NC10): 9776–9782. http://dx.doi.org/10.1029/JC086iC10p09776.

Warren JM and Hirth G (2006) Grain size sensitive deformation mechanisms in naturally deformed peridotites. *Earth and Planetary Science Letters* 248(1–2): 438–450. http://dx.doi.org/10.1016/j.epsl.2006.06.006.

Weatherley SM and Katz RF (2012) Melting and channelized magmatic flow in chemically heterogeneous, upwelling mantle. *Geochemistry, Geophysics, Geosystems* 13. http://dx.doi.org/10.1029/2011GC003989.

Weertman J and Weertman JR (1975) High-temperature creep of rock and mantle viscosity. *Annual Review of Earth and Planetary Sciences* 3: 293–315.

Weinstein SA (1996) Thermal convection in a cylindrical annulus with a non-Newtonian outer surface. *Pure and Applied Geophysics* 146: 551–572.

Weinstein SA and Christensen U (1991) Convection planforms in a fluid with a temperature-dependent viscosity beneath a stress-free upper boundary. *Geophysical Research Letters* 18: 2035–2038.

Weinstein SA and Olson PL (1992) Thermal-convection with non-Newtonian plates. *Geophysical Journal International* 111: 515–530.

Williams Q and Garnero EJ (1996) Seismic evidence for partial melt at the base of Earth's mantle. *Science* 273: 1528–1530.

Woods LC (1975) *The Thermodynamics of Fluid Systems*. Oxford: Clarendon Press.

Yuen D, Sabadini RCA, and Gasperini P (1986) On transient rheology and glacial isostasy. *Journal of Geophysical Research* 91: 1420–1438.

Zhang SX and Christensen U (1993) Some effects of lateral viscosity variations on geoid and surface velocities induced by density anomalies in the mantle. *Geophysical Journal International* 114: 531–547.

Zhang SX and Yuen DA (1996) Various influences on plumes and dynamics in time-dependent, compressible mantle convection in 3-D spherical shell. *Physics of the Earth and Planetary Interiors* 94: 241–267.

Zhong SJ, Gurnis M, and Moresi L (1996) Free surface formulation of mantle convection, part 1: Basic theory and implication to plumes. *Geophysical Journal International* 127: 708–718.

Zhong SJ, McNamara A, Tan E, Moresi L, and Gurnis M (2008) A benchmark study on mantle convection in a 3-D spherical shell using citcoms. *Geochemistry, Geophysics, Geosystems* 9: Q07005. 11063–11082. http://dx.doi.org/10.1029/2008GC002048.

Zhong SJ, Zuber MT, Moresi L, and Gurnis M (2000) Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *Journal of Geophysical Research* 105: 11063–11082.

This page intentionally left blank

7.03 Laboratory Studies of Mantle Convection

A Davaille, CNRS/University of Paris-Sud, Orsay, France
A Limare, Sorbonne Paris Cité/Université Paris-Diderot/CNRS, Paris, France

© 2015 Elsevier B.V. All rights reserved.

7.03.1	**Introduction**	74
7.03.2	**Experimental Setups and Fluids**	76
7.03.2.1	Designing an Experiment: Scaling	76
7.03.2.2	Experimental Fluids	76
7.03.2.3	Experimental Setup for Convection	78
7.03.2.3.1	Horizontal pattern visualization	78
7.03.2.3.2	Heat flux determination	78
7.03.2.3.3	Hybrid solution at high *Pr* and *Ra*	78
7.03.2.3.4	Moving boundaries and free-slip boundaries	78
7.03.2.3.5	Changing gravity: centrifuge and microgravity	79
7.03.3	**Measurements and Visualization Techniques**	80
7.03.3.1	Patterns	80
7.03.3.1.1	Dye	80
7.03.3.1.2	Shadowgraph and schlieren	80
7.03.3.2	Temperature and Heat Flow Measurements	83
7.03.3.2.1	Local temperature measurements	83
7.03.3.2.2	Deflection of a light beam: 1-D profile	85
7.03.3.2.3	Interferometry: 2-D and 3-D	85
7.03.3.2.4	3-D fields: thermochromic liquid crystals and fluorescence	87
7.03.3.2.5	Heat flow measurements	88
7.03.3.3	Composition	89
7.03.3.4	Fluid Motions	89
7.03.3.4.1	Local measurements	89
7.03.3.4.2	2-D and 3-D field measurements	90
7.03.3.4.3	Stress and strain rate	90
7.03.3.5	Surface Topography	90
7.03.4	**Rayleigh–Taylor Instabilities**	92
7.03.5	**Simple Rayleigh–Bénard Convection Studies**	94
7.03.5.1	Convection at Relatively Low *Ra* ($Ra_c \leq Ra \leq 10^6$)	94
7.03.5.2	Patterns and Defects	95
7.03.5.2.1	Generation of an initial prescribed pattern	95
7.03.5.2.2	Patterns for $Pr \geq 100$	95
7.03.5.2.3	Patterns for $Pr < 100$	96
7.03.5.3	Thermal Convection at High Rayleigh Numbers ($10^5 \leq Ra$)	97
7.03.5.3.1	TBL instabilities	97
7.03.5.3.2	Soft and hard turbulence at low *Pr*	98
7.03.5.3.3	Large-scale (cellular) circulation and plumes at high *Pr*	98
7.03.5.4	Convective Regime in an Isoviscous Mantle	99
7.03.6	**Temperature-Dependent Viscosity**	101
7.03.6.1	Rigid Boundaries	101
7.03.6.1.1	Onset of convection	101
7.03.6.1.2	Patterns and regimes	101
7.03.6.1.3	Characteristics of the stagnant lid regime	103
7.03.6.2	Stagnant Lid Regime and Lithosphere Cooling	104
7.03.6.3	Plate Tectonics in the Laboratory?	106
7.03.6.3.1	Free slip boundaries	106
7.03.6.3.2	Several scales of convection	106
7.03.6.3.3	'True' plate tectonics	106
7.03.7	**Complications: Internal Heating, Continents, and Moving Plates**	107
7.03.7.1	Internal Heating	107
7.03.7.1.1	Purely internally heated fluid	107
7.03.7.1.2	Bottom-heated and internally heated fluid	107
7.03.7.2	Inhomogeneous Cooling/Continents	108

7.03.7.3 Moving Top Boundary 109
7.03.8 Close-Up on Plumes: Plumes from a Point Source of Buoyancy 110
7.03.8.1 Compositionally Buoyant Plumes 110
7.03.8.2 Thermal Plumes 111
7.03.8.2.1 Thermals 112
7.03.8.2.2 The stem 112
7.03.8.2.3 Thermal starting plumes 114
7.03.8.3 Plumes in Non-Newtonian Fluids 115
7.03.8.4 Interaction of Plumes with a Large-Scale Flow 115
7.03.8.5 Plume–Lithosphere Interactions 116
7.03.8.5.1 Stagnant lithosphere 116
7.03.8.5.2 Moving or rifting lithosphere 116
7.03.9 Inhomogeneous Fluids: Mixing and Thermochemical Convection 117
7.03.9.1 Mixing of Passive Tracers 117
7.03.9.2 Convection in an Initially Stratified Fluid 118
7.03.9.2.1 Regime diagrams 118
7.03.9.2.2 Whole-layer regime and unstable doming 119
7.03.9.2.3 Stratified regime: anchored hot spots, bumps, ridges, and piles 121
7.03.9.2.4 Relevance of the experiments to the deep mantle: pressure dependence of B 122
7.03.9.2.5 Stability of the continental lithosphere 122
7.03.9.2.6 Evolution of an initially stratified mantle and continental growth 123
7.03.9.2.7 Diversity of mantle upwellings 123
7.03.9.3 Entrainment and Mixing 124
7.03.9.3.1 Entrainment from an interface 124
7.03.9.3.2 Stirring 124
7.03.9.3.3 Experimental limitation: species diffusion 125
7.03.9.4 Interaction of a Plume with a Density and/or Viscosity Interface 125
7.03.10 Mid-Ocean Ridges and Wax Tectonics 126
7.03.10.1 Can a Buoyant Mantle Upwelling Generate Sufficient Stress to Rupture a Brittle Crust? 126
7.03.10.2 Morphology of Ridges 126
7.03.11 Subduction-Related Experiments 127
7.03.11.1 Ingredients for Subduction 128
7.03.11.2 The Surface Story and the Initiation of Subduction 130
7.03.11.3 Mantle Flow Induced by a Cold Rigid Slab 131
7.03.11.3.1 Generation of seismic anisotropy 131
7.03.11.3.2 Mantle flow and thermal evolution of the slab 131
7.03.11.4 Coupled Mantle–Slab Fluid System 131
7.03.11.4.1 Fixed rollback velocity 131
7.03.11.4.2 Free rollback velocity 132
7.03.12 Conclusions 134
Acknowledgments 135
References 135

7.03.1 Introduction

Because laboratory experiments are crucial for exploring new physics and testing theories, they have long played a central role in investigations of thermal convection and mantle dynamics. This chapter is devoted to laboratory experiments considered as tools for understanding the physics that governs mantle dynamics. We shall review both the techniques employed to run well-controlled experiments and acquire quantitative data and the results obtained.

Investigations of mantle dynamics and of thermal convection have long been closely intertwined. Indeed, the mantle is cooled from above, heated from within by radioactive elements, and heated from below by the core, which loses its heat through the mantle (see Jaupart et al., Chapter 7.06). The emergence of mantle convection models was dictated by the failure of static, conductive, and/or radiative thermal history models to account for the mantle temperature regime, the Earth's energy budget, and the Earth's lateral surface motions. In the late 1930s, Arthur Holmes, among others, hypothesized that thermal convection in the Earth's mantle provided the necessary force to drive continental motions (Holmes, 1931). Convection, which transports heat by material advection, is the only other physical mechanism capable of explaining these observations. The force driving advection is gravity, whereby material lighter than its environment rises, while denser material sinks. Such density anomalies can be produced by differences in composition and/or temperature.

In the simple configuration of a plane layer (Figure 1), the former will give rise to 'Rayleigh–Taylor' instabilities, while the latter will generate 'Rayleigh–Bénard' instabilities.

Rayleigh–Bénard convection develops when a plane layer of fluid is heated from below and cooled from above (Figure 1). It has been identified as a major feature of the dynamics of the oceans, the atmosphere, and the interior of stars and planets. First identified by Count Rumford (1798), the phenomenon was observed several times in the nineteenth century (e.g., Thomson, 1882). However, it was the carefully controlled and quantitative laboratory experiments of Bénard (1900, 1901) that focused the interest of the scientific community on the problem. Bénard studied the patterns of convection developing in thin layers with a free upper surface. He was interested in the influence of viscosity on the pattern and used several fluids, including spermaceti and paraffin. He determined quantitatively the characteristic length scales of the patterns, the deformation of the interface, and the direction of flow within the fluid. Although we now know that the beautiful hexagonal patterns he observed (Figure 1(c)) were due to the temperature dependence of surface tension (Pearson, 1958), it was these experiments that motivated Lord Rayleigh to apply hydrodynamic stability theory to thermal convection in the absence of surface tension (Rayleigh, 1916). When the thermal convection experiments were carried out correctly, they were found to be very well predicted by the Rayleigh theory (e.g., Schmidt and Milverton, 1935; Silveston, 1958). For more on the history of these investigations, see Chapter 7.01 of this volume.

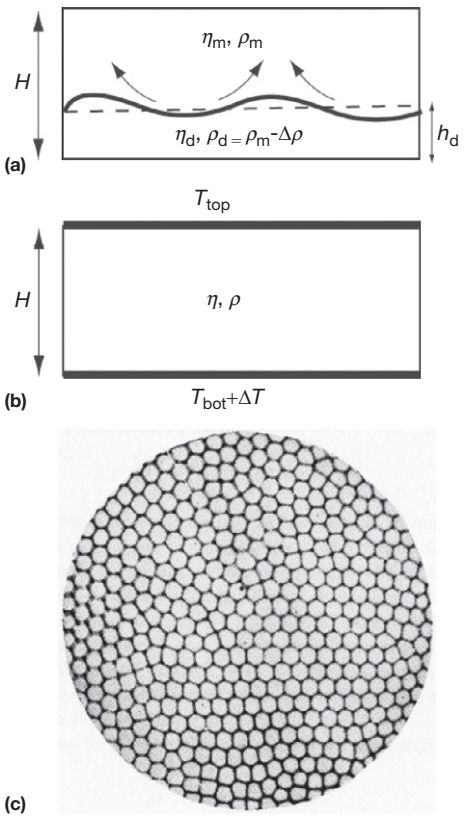


Figure 1 (a) Rayleigh–Taylor instabilities develop when a layer of lighter fluid is introduced under a layer of denser fluid. (b) Rayleigh–Bénard instabilities develop when a layer of fluid is heated from below and cooled from above. (c) Bénard cells in spermaceti. Reproduced from Bénard H (1901) Les tourbillons cellulaires dans une nappe liquide transportant la chaleur par convection en régime permanent. *Annales de Chimie Physique* 23: 62–144.

Thermal convection in an isoviscous fluid is characterized by two parameters. The Rayleigh number $Ra(H, \Delta T)$ compares the driving thermal buoyancy forces to the resisting effects of thermal diffusion and viscous dissipation across the whole system:

$$Ra(H,\Delta T) = \frac{\alpha g \Delta T H^3}{\kappa \nu} \quad [1]$$

where H is the depth of the layer, ΔT the temperature difference applied across it, g the gravitational acceleration, α the thermal expansivity, κ the thermal diffusivity, and ν the kinematic viscosity. Convection starts when Ra exceeds a critical value (Rayleigh, 1916; Ricard, Chapter 7.02; Ribe, Chapter 7.04) and exhibits a sequence of transitions toward chaos as Ra increases. The second parameter is the Prandtl number, Pr, the ratio of the diffusivity of momentum and that of heat:

$$Pr = \nu/\kappa \quad [2]$$

When $Pr \gg 1$, the fluid motion stops as soon as the buoyancy source (due here to temperature differences) disappears, that is, inertial effects are negligible compared to viscous effects. This is the case for the Earth's mantle, where $Pr > 10^{23}$. Numerous theoretical, laboratory, and numerical studies in the last 50 years have been devoted to characterizing thermal convection as a function of Ra and Pr. Laboratory experiments have proved especially useful for determining patterns and characterizing the high-Ra regime. For high-Pr fluids, isoviscous convection has been studied for Ra up to 10^9, which includes the range of values estimated for the mantle (10^6–10^8).

However, mantle dynamics is much more complicated than isoviscous convection, in particular due to the complex rheology of mantle material (Ricard, Chapter 7.02). For example, the generation of plate tectonics and its coexistence with hot spots are not reproduced in isoviscous fluids. Studies have therefore focused on progressively more complicated systems, taking either a global view (e.g., convection with temperature-dependent rheology or internal heating) or a more local view to study some particular mantle feature (such as plumes or subduction). Analogue laboratory experiments have been extensively used because of four advantages: (a) Since they let nature solve the equations, they can explore new phenomena for which such equations do not yet exist. (b) They can also explore ranges of parameters, or geometries, where the equations are too nonlinear to be solved analytically or numerically. (c) They are inherently three-dimensional (3-D). (d) They can usually be run in the appropriate range of mantle parameters (which is not the case for the atmosphere or the oceans).

This chapter is organized as follows: Sections 7.03.2 and 7.03.3 are devoted to experimental setups, fluids, measurements, and visualization methods. Mantle dynamics on geologic timescales is dominated by 'fluid' behavior, so that we can generally use liquids around room temperature, and fluid mechanics techniques. With the development of computer power and lasers, it has become possible in the last years to measure the temperature, velocity, concentration, and deformation fields in

experimental tanks. Sections 7.03.4–7.03.7 focus on gravitational instabilities (Rayleigh–Taylor instabilities and Rayleigh–Bénard convection). Sections 7.03.8–7.03.11 will describe the laboratory experiments related to more specific mantle features – plumes, mixing, accretion, and subduction – which are also described more fully in other chapters of this volume.

7.03.2 Experimental Setups and Fluids

7.03.2.1 Designing an Experiment: Scaling

The goal of any fluid mechanics modeling is to determine scaling laws or functional relations that link certain parameters of interest and the various other parameters on which they depend. These scaling laws then make it possible to predict the behavior of similar systems, such as the mantle in geodynamics. Hence, there is no question of building a miniature Earth in the laboratory. A phenomenon has to be selected, and a simplified laboratory model is then constructed, where only a few parameters can vary in a controlled manner. Each individual experiment aims at describing, through quantitative measurements, the behavior of the system for a given set of control parameters. By varying systematically the values of these parameters, a database is constituted. Scaling laws derived from fundamental physical principles can then be tested against the experimental data.

However, the results of a laboratory experiment will be applicable to other natural systems, such as the mantle, only if the dynamic similarity between the scaled-down system (laboratory and computer) and the natural system is respected. Dynamic similarity can be viewed as a generalization of the concept of geometric similarity (see Ribe, Chapter 7.04) and requires (a) similar boundary conditions (mechanical, thermal, and geometry), (b) similar rheological laws (in other words, the mechanical equation of state that relates differential flow stress to strain rate should differ only in the proportionality constant (e.g., Weijermars and Schmeling, 1986)), and (c) similar balances between the different forces and operative physical effects (e.g., Ribe and Davaille, 2013). *Ra* and *Pr* reflect such balances. Dynamic similarity gives birth to a set of dimensionless parameters. When the governing equations are known, they can be nondimensionalized to make the relevant dimensionless parameters appear explicitly in the equations and/or the initial and boundary conditions. When the equations are not known, one may use the Buckingham Π theorem, which is a consequence of the fundamental principle that the validity of a physical law cannot depend on the units in which it is expressed. According to this theorem, if a given experiment is described by N-dimensional parameters of which M have independent physical dimensions, then the experiment can be completely described by $N-M$-independent nondimensionless combinations of the dimensional parameters (see Ribe, Chapter 7.04).

Besides *Ra*, *Pr*, and ratios that characterize the variations of physical properties such as viscosity within the experimental tank, other important dimensionless numbers for geodynamics include the following:

- the Reynolds number *Re*, which compares the advection of momentum by the fluid motion and the viscous diffusion of momentum:

$$Re = \frac{UH}{\nu} \quad [3]$$

where U is the characteristic velocity in the system. For the mantle, $Re \sim 10^{-19}$.

- the Péclet number, *Pe*, when the flow is forced by a boundary velocity U (such as a plate velocity). *Pe* compares the heat transport by advection and the heat transport by conduction:

$$Pe = \frac{Ud}{\kappa} \quad [4]$$

where d is, for example, the thickness of the plate.

All systems with the same dimensionless parameter (say, *Ra*) will behave in the same way, irrespective of their size. However, the timescale and/or the distance over which the phenomenon occurs will depend on the system size. This is how convection in the laboratory on a scale of hours can be analogous to convection in the mantle over geologic times. Dynamic similarity and scaling analysis are therefore essential to analogue laboratory modeling. Their principles and techniques are discussed in Chapter 7.04.

7.03.2.2 Experimental Fluids

Well-controlled experiments require a precise knowledge of the physical properties of the fluids used. Thermal conductivity and specific heat do not change much from one batch to another, and one can generally rely on the values given by the manufacturer. But it is advisable to measure density (and thermal expansion if needed) and viscosity prior to any new experiments. Moreover, because of the temperature dependence of viscosity, experiments should preferably be run in a temperature-controlled laboratory. This will also ensure that electronic measurements do not drift with time.

Except when focusing on lithospheric processes such as accretion or subduction, laboratory experiments have mostly assumed that mantle material flows like a Newtonian fluid, with a linear relation between stress and strain rate. This is especially true for convection experiments, because non-Newtonian fluids are usually more difficult to characterize and to handle.

In the mantle, $Pr \sim 10^{24}$. This is not possible to obtain in the laboratory, but because the dynamics becomes independent of *Pr* for $Pr > 100$ (Krishnamurti, 1970a,b; Whitehead, 1976; Whitehead and Parsons, 1978; see Section 7.03.5), the use of fluids with $Pr > 100$ is adequate to ensure the dominance of viscous over inertial effects.

Another consideration is the temperature dependence of viscosity. Solid-state creep is thermally activated, and therefore, the mantle material has a highly temperature-dependent viscosity (e.g., Kohlstedt, vol. 2). The viscosity ratio across the lithosphere where the temperature difference is the highest reaches 10^7. A number of experiments have investigated in detail the influence of a strongly temperature-dependent viscosity on convection. The viscosity of liquids always depends on temperature. The silicone oils have the smallest, with a change by no more than a factor of 3 over 50 °C. Sugar syrups have a much stronger temperature dependence, with golden syrup (from Tate and Lyle) the winner, showing a 10^6 viscosity change between 60 and −20 °C (Figure 2).

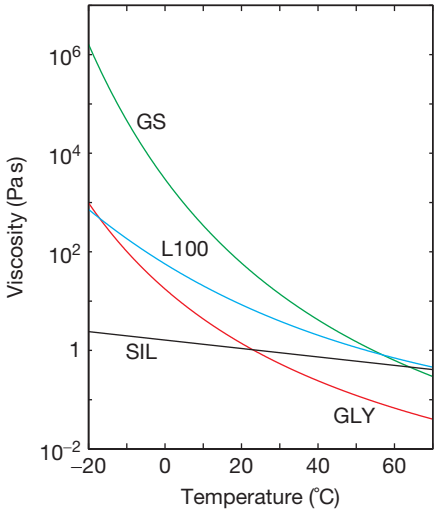


Figure 2 Viscosity as a function of temperature. SIL, silicone oil 47 V1000; GLY, glycerol; GS, golden syrup; L100, polybutene.

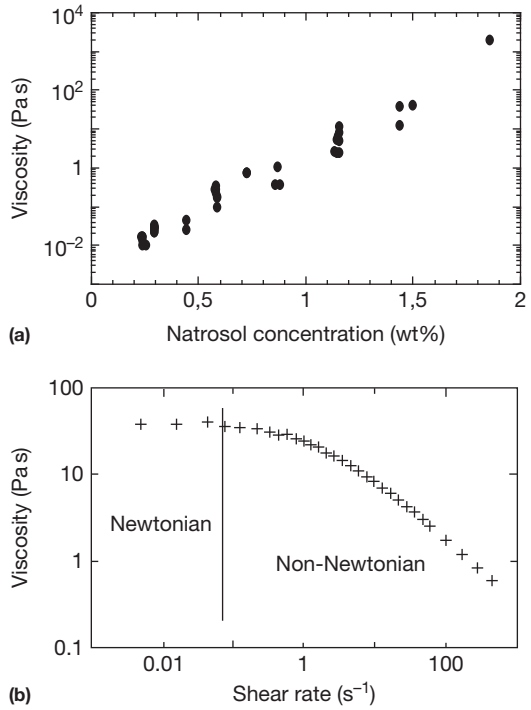


Figure 3 Natrosol aqueous solutions. (a) Viscosity as a function of shear rate (measured with a Haake rotational viscometer) for 1.4% wt Natrosol. There is a Newtonian plateau for low shear rates. (b) Viscosity as a function of the Natrosol concentration. Reproduced from Davaille A (1999a) Two-layer thermal convection in viscous fluids. *Journal of Fluid Mechanics* 379: 223–253.

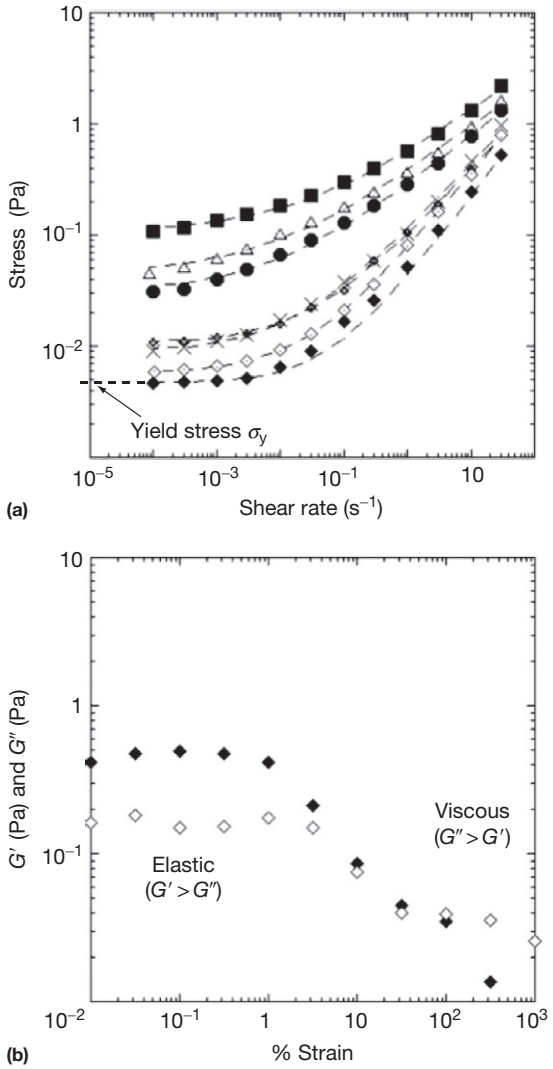


Figure 4 Rheological properties of Carbopol gels at $T = 293$ °K. (a) Steady-state rheometry: Shear stress σ as a function of strain rate ($\dot{\epsilon}$). (Black diamonds: $\sigma_Y = 0.0047$ Pa, diamonds: $\sigma_Y = 0.006$ Pa, plus: $\sigma_Y = 0.01$ Pa, crosses: $\sigma_Y = 0.009$ Pa, black circles: $\sigma_Y = 0.031$ Pa, triangles: $\sigma_Y = 0.045$ Pa, black squares: $\sigma_Y = 0.104$ Pa, dashed lines: Herschel–Bulkley model whereby $\sigma = \sigma_Y + K_v \dot{\epsilon}^n$, K_v is the consistency, and n is the shear-thinning exponent). (b) Oscillatory rheometry: Variation of the elastic G' (black diamonds) and loss G'' (white diamonds) moduli as a function of the strain for a Carbopol gel with $\sigma_Y = 0.006$ Pa and a frequency $f = 0.1$ Hz. Adapted from Darbouli M, Metivier C, Piau J.-M., Magnin A, and Abdelali A (2013) Rayleigh–Bénard convection for viscoplastic fluids. *Physics of Fluids* 25: 023101.

Silicone oils are available in different grades (with viscosity at 20 °C between 10 mP s and 10^3 Pa s) and should be used when negligible temperature-related viscosity variations are required in an experiment. Thickeners in aqueous solutions have also been used (e.g., Davaille, 1999a,b; Namiki, 2003; Tait and Jaupart, 1989). By adding <2% wt of thickener (e.g., hydroxyethylcellulose, trade name Natrosol), the viscosity of water can be multiplied by 10^7 (Figure 3). Although the resulting fluid is shear thinning (Figure 3(b)), there is a Newtonian plateau at the low shear rates that typically obtain in convection laboratory experiments.

A few experiments have lately studied the influence of nonlinear rheology on thermal convection (Balmforth and Rust, 2009; Darbouli et al., 2013; Davaille et al., 2013; Kebiche et al., 2013; Massmeyer et al., 2013), using aqueous mixtures of a polymeric microgel, Carbopol. These fluids exhibit a viscoelastoplastic behavior (Figure 4) whereby they are shear thinning when the local stress exceeds a critical stress σ_Y (the 'yield' stress) and deform elastically otherwise (e.g., Piau, 2007).

7.03.2.3 Experimental Setup for Convection

A thermal convection experiment should minimize heat losses to allow reliable heat flux determination and should also allow good flow visualization (pattern and thermal structure). Unfortunately, these two requirements cannot be perfectly met at the same time. We shall therefore present three commonly used experimental setups.

7.03.2.3.1 Horizontal pattern visualization

The convecting fluid is bounded between two glass plates; the bottom is heated from below by hot water flushing along its outer surface, while the top is cooled by cold water flushed along the top, as shown on **Figure 5(a)** (e.g., Busse and Whitehead, 1971; Weinstein and Christensen, 1991; Weinstein and Olson, 1990; White, 1988; Whitehead and Parsons, 1978). An alternative is to use as the bottom heater a metal plate whose upper surface is polished to a mirror finish (e.g., Chen and Whitehead, 1968; Heutmaker and Gollub, 1987). The horizontal heat exchangers should be carefully levelled since imperfect horizontal alignment modifies convection patterns and heat transfer (e.g., Chilla et al., 2004; Namiki and Kurita, 2002). Precautions also have to be taken to eliminate lateral inhomogeneities due to the sidewalls. In Busse and Whitehead's setup (**Figure 5(a)**), the area where observations on convection were made was bounded on the sides by walls of 2 inches thick polyvinyl chloride, whose thermal conductivity is close to that of the working fluid (silicone oil). This provided a working area 80×80 cm. Outside these walls was another convecting region $\sim$20 cm in width so that the temperature gradients on both sides of the walls were similar. Visualization was done using the shadowgraph technique (see **Section 7.03.3**), whereby collimated light (i.e., with parallel rays) is directed vertically up through the tank (**Figure 5(a)**). In the convecting fluid, the index of refraction of light depends on temperature such that light rays diverge away from hot regions and converge toward colder regions. The projection of the resulting rays onto a white screen thus show hot zones as dark shadows and cold zones as concentration of brightness.

7.03.2.3.2 Heat flux determination

The dependence of the global heat transport on convection characteristics is one of the fundamental questions of Rayleigh–Bénard studies (see **Section 7.03.5**). The most accurate measurement of the global heat flow is obtained through the measurement of the electric power needed to heat the bottom plate (e.g., Brown et al., 2005; Funfschilling et al., 2005; Giannandrea and Christensen, 1993; Malkus, 1954; Schmidt and Milverton, 1935; Silveston, 1958). It requires special design of the experimental tank in order to minimize the heat losses from the heating plate to the bottom (e.g., **Figure 5(b)**). A second experimental problem is the influence of the sidewalls on the heat transport in the fluid. This problem is significantly reduced when the experimental tank aspect ratio is large (e.g., Giannandrea and Christensen, 1993) and by choosing a wall of relatively small thermal conductivity (e.g., Plexiglas) compared to that of the fluid, plus another insulation in low conductivity material such as polystyrene foam. Even with that, side losses must be estimated, and models have been derived to correct for them (e.g., Ahlers, 2001). A third problem is the effect of the finite conductivity of the top and bottom plates on the heat transport by the fluid (e.g., Chilla et al., 2004; Verzicco, 2004): the less conductive the plates, the more the heat transport in the fluid is diminished, and the effect is a function of *Ra*. A correction has been proposed, which fits well the data for turbulent convection (Brown et al., 2005; Verzicco, 2004). In practice, the best conductor available is copper ($k = 319\ \mathrm{W\ m^{-1}\ K^{-1}}$), followed by aluminum ($k = 161\ \mathrm{W\ m^{-1}\ K^{-1}}$). So good heat flux determination precludes the visualization of the planform because of the metal plates and also visualization from the side because of the insulation.

7.03.2.3.3 Hybrid solution at high Pr and Ra

It is of course necessary to correlate heat flux measurements with the geometry of convection. Moreover, modern visualization techniques can give *in situ* measurements of the velocity, temperature, and concentration fields (see **Section 7.03.3**), but they require transparent sides. Therefore, a number of studies have used a lighter setup (**Figure 5(c)**; e.g., Davaille and Jaupart, 1993; Jaupart and Brandeis, 1986; Olson, 1984; Weeraratne and Manga, 1998). The bottom and top heat exchangers are made of copper or aluminum and are regulated either by electric heating or by circulation of thermostated water. The tank sides are made of Plexiglas or glass. The whole tank is insulated with Styrofoam, but the insulation on the sides can be removed for visualization. In this configuration, it is advisable to run the experiments in conditions such that the mean temperature inside the experimental cell is close to ambient temperature, to minimize heat losses. Last, to obtain high Rayleigh numbers (10^6–10^9) in high-*Pr* viscous fluids, we need fluid thicknesses between at least 10 and 50 cm. It is then no longer possible to have large aspect ratios. It therefore will always be a possibility that the flow will be influenced by the mechanical boundary conditions (zero velocity) at the sidewalls. We shall discuss this question in more detail later.

7.03.2.3.4 Moving boundaries and free-slip boundaries

To impose a moving upper boundary, a mylar sheet (**Figure 5(a)**) is usually introduced just below the upper plate and slowly driven by a stepping motor (e.g., Jellinek et al., 2003; Kincaid et al., 1995, 1996; Richter and Parsons, 1975). Because the mantle is bounded by the liquid core at the bottom, and oceans or atmosphere at the top, its mechanical outer boundary conditions are free-slip. Moreover, analytic and numerical models usually are best resolved with free-slip boundary conditions. To obtain the latter in the laboratory, thin layers of a fluid much less viscous (at least a 1000 times) than the working fluid must be introduced between the heat exchangers and the high-*Pr* convecting fluid. For a free-slip bottom condition, thin layers of mercury (e.g., Solomon and Gollub, 1991) or salted water (e.g., Jellinek and Manga, 2002, 2004; Jellinek et al., 2002) may be used, while a thin layer of conductive oil may be used to obtain a free-slip upper surface (e.g., Giannandrea and Christensen, 1993). However, since these layers have a finite conductivity, they will modify the thermal boundary conditions as discussed earlier.

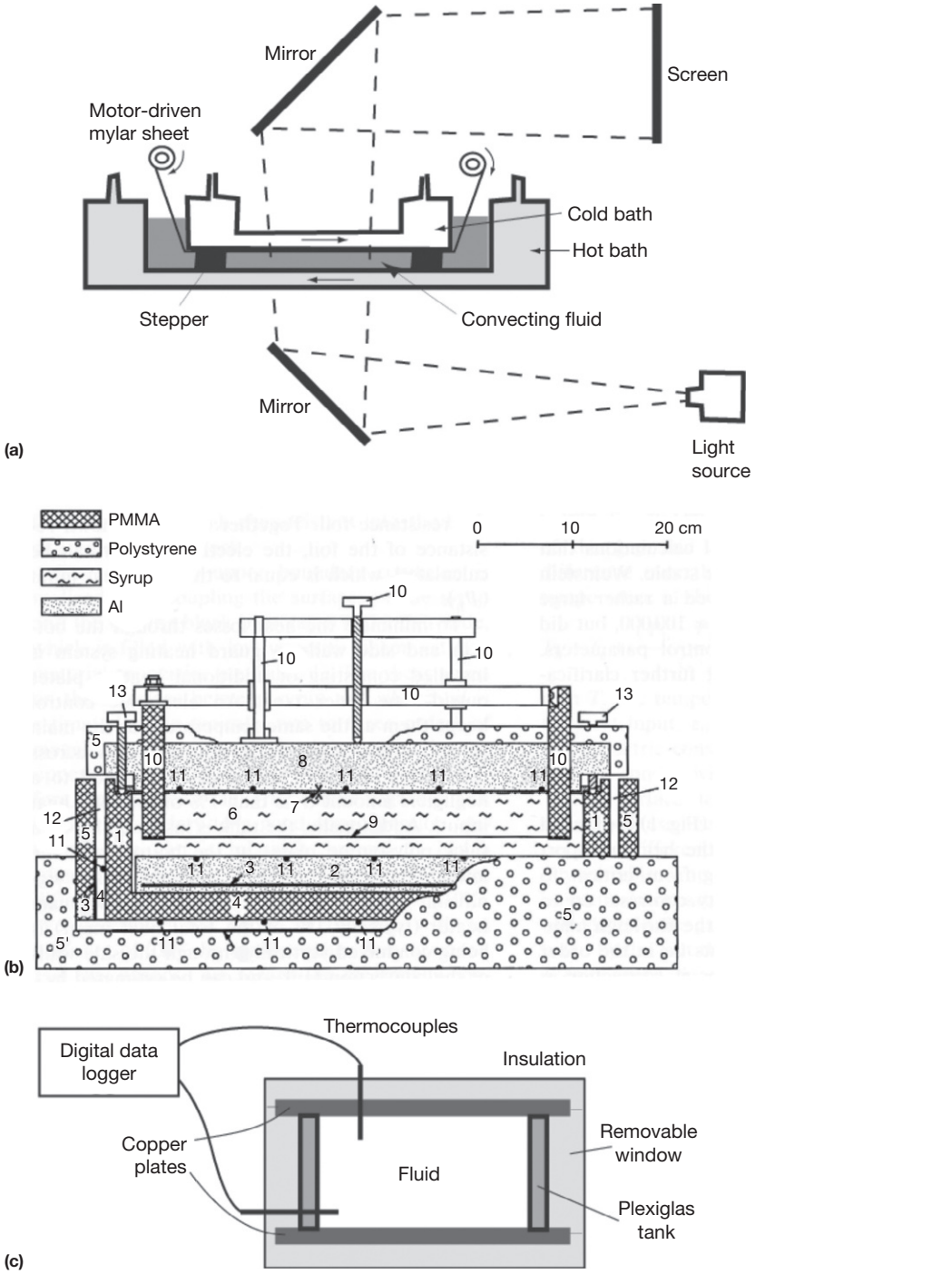


Figure 5 (a) Experimental setup from Richter and Parsons (1975) (adapted from Busse FH and Whitehead JA (1971) Instabilities of convection rolls in high Prandtl number fluid. *Journal of Fluid Mechanics* 47: 305–320). Thermostated water is flowed in the top and bottom glass assemblages. A mylar sheet can be introduced just underneath the top cold plate and driven by a motor at constant speed. (b) Convection apparatus from Giannandrea and Christensen (1993). 1, convection tank; 2, heating plate; 3, heating foil; 4, guard heaters; 5, thermal insulation; 6, working fluid; 7, oil-filled gap; 8, cooling block; 9, nickel wire; 10, device for vertical displacement of the Ni-wire; 11, thermocouples; 12, air gap; 13, adjusting screw. c) "hybrid" experimental set up. The insulation windows can be removed for visualization.

7.03.2.3.5 Changing gravity: centrifuge and microgravity

Some experiments have been run in centrifuges (e.g., Nataf et al., 1984; Ramberg, 1972; Tentler, 2003a,b, 2007; Weijermars, 1988), where the model is subjected to a centrifugal force able to produce accelerations up to 20 000g. This approach decreases the experimental time required and allows the use of very viscous materials (with $Pr > 10^6$) at Rayleigh numbers up to 6×10^5. A detailed description of the

experimental setup is given by Nataf et al. (1984) and Weijermars (1988). However, this technique is heavy to implement. With the apparent confirmation (in part because of the early centrifuge experiments) that convective dynamics becomes independent of P when $Pr > 100$, the technique has not been used extensively.

Most of the experiments in the laboratory are carried out in a Cartesian geometry. However, given the Earth's geometry, experiments in spherical shells and radial force fields are wanted. Using dielectric convection in a spherical conductor generates a highly regular central force field that decays rapidly with radius, resulting in an effective gravity $g_e \sim 1/r^5$. A new experiment, 'GeoFlow,' has therefore been built on this principle (Egbers et al., 2003). However, the g_e rapid decay and the very high voltages required limit the attainable gravity. Hence, to simulate a central force everywhere in the fluid, 'GeoFlow' is running in the microgravity environment of the International Space Station (**Figure 6**; Futterer et al., 2010, 2012, 2013).

7.03.3 Measurements and Visualization Techniques

7.03.3.1 Patterns

Laboratory experiments have been used extensively to determine convective and/or mixing patterns. One then seeks quantitative information on the morphology, wavelength, and evolution (plume ascent rate, blob stretching, etc.) of a particular feature in the experimental fluid.

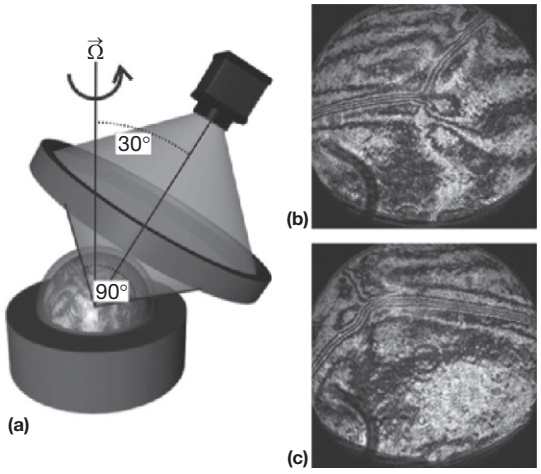


Figure 6 GeoFlow experiment. (a) Experimental setup; (b) and (c) Interferograms of the north pole of the sphere using Wollaston shearing interferometry for $Ra = 5 \times 10^4$ (top) and $Ra = 6 \times 10^4$ (bottom). Adapted from Futterer B, Egbers Ch, Koch S, Dahley N, and Jehring L (2010) First identification of sub- and supercritical convection patterns from GeoFlow, the geophysical flow simulation experiment integrated in Fluid Science Laboratory. *Acta Astronautica* 66: 193–200; Futterer B, Dahley N, Koch S, Scurtu N, and Egbers Ch (2012) From isoviscous convective experiment GeoFlow I to temperature-dependent viscosity in GeoFlow II – Fluid physics experiments on-board ISS for the capture of convection phenomena in Earth's outer core and mantle. *Acta Astronautica* 71: 11–19.

7.03.3.1.1 Dye

7.03.3.1.1.1 One-shot visualization

Patterns and motions in two-fluid experiments such as compositional plumes (**Figure 7**; Olson and Singer, 1985; **Section 7.03.6**), Rayleigh–Taylor instabilities (e.g., Whitehead and Luther, 1975; see **Section 7.03.4**), mixing (e.g., Ottino, 1989; see **Section 7.03.9**), or two-layer convection (e.g., Davaille, 1999a,b; Fourel et al., 2013; Kumagai et al., 2008; Olson, 1984; Olson and Kincaid, 1991; see **Section 7.03.9**) are easily observed by the addition of dye to one of the transparent fluids. Most commonly used dyes are fluorescein, rhodamine, and supermarket food dye. The latter is very easily visualized through illumination with a white light (**Figure 7**), while the former are most spectacular using laser sheets (e.g., Fountain et al., 2000; Tsinober et al., 1983). In mixing experiments, the dye can be continuously injected by syringes (e.g., Fountain et al., 2000; Kerr and Mériaux, 2004). In all cases, the dye quantity is so small that it does not change the physical properties of the fluids.

The fluid initially at rest can also be marked by dye streaks that allow subsequent fluid motions to be recorded by their distortion. Griffiths (1986a,b) visualized the motions of initially cold fluid due to the passage of a thermal by injecting before the run at various heights a number of horizontal lines of the same fluid containing a small concentration of dye (**Figure 8(a)**). Using very viscous putties, Weijermars (1988, 1989) imprinted black grids on some cross sections prior to experiments in a centrifuge, which recorded the subsequent deformation (**Figure 8(b)**).

7.03.3.1.1.2 Electrochemical technique

In water and aqueous solutions, an electrochemical technique using thymol blue, a pH indicator, can be employed (Baker, 1966). Thymol blue is either blue or yellow-orange depending upon whether the pH of the solution is greater or less than 8. Approximately 0.01% by weight of thymol blue is added to the water, and the solution is titrated to the end point with sodium hydroxide. Then, by drop addition of hydrochloric acid, the solution is made orange in color. When a small DC voltage (~10–20 V) from a dry-cell source is impressed between a pair of electrodes situated within such a fluid, H^+ ions are removed from solution at the negative electrode, with a corresponding change in color from orange to blue. The dye thus created is neutrally buoyant and faithfully follows the motion of the fluid. Sparrow et al. (1970) used the bottom plate of the experimental cell as the negative electrode itself, while the positive electrode was a large copper sheet situated adjacent to the fluid but well removed from the bottom plate. They hereby obtained the first photographs of thermals rising from a heated horizontal surface (**Figure 9**). Haramina and Tilgner (2004) recently used the same method to image coherent structures in boundary layers of Rayleigh–Bénard convection at very high Rayleigh numbers.

7.03.3.1.2 Shadowgraph and schlieren

Since the refractive index of a fluid depends on temperature and composition (e.g., salt or sugar content), convective features in a transparent medium will alter the refractive index distribution, also called schlieren (a German term sometimes spelled shlieren in the literature). From the deformation of the

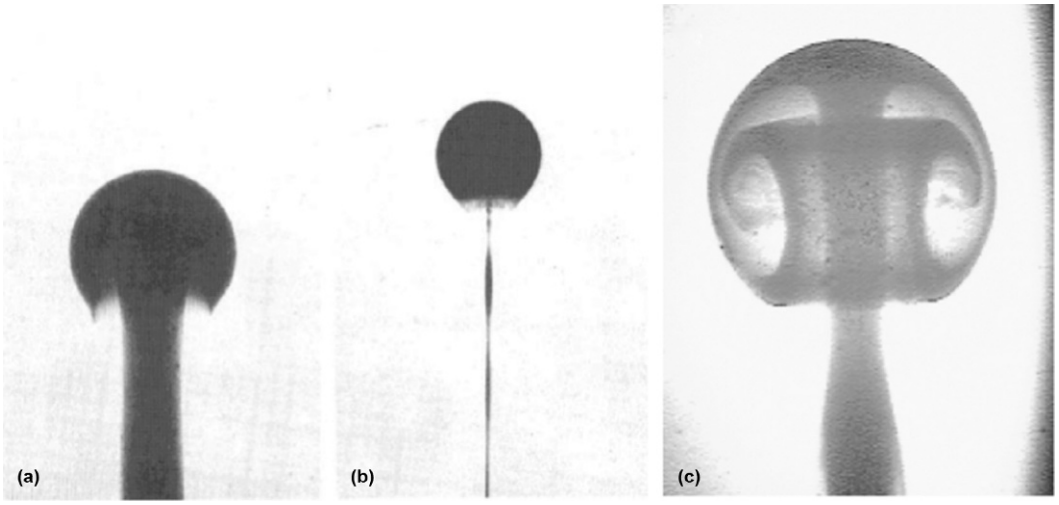


Figure 7 Morphology of starting plumes when the dyed plume material is injected into a transparent medium. (a and b) Purely compositional plumes (reproduced from Olson P, Singer H (1985) Creeping plumes. *Journal of Fluid Mechanics* 158: 511531): (a) Diapiric plume, more viscous than its surroundings. (b) Cavity plume, less viscous than its surroundings. (c) Starting thermal plume: the injected dyed material is hotter and less viscous. Reproduced from Laudenbach N and Christensen UR (2001) An optical method for measuring temperature in laboratory models of mantle plumes. *Geophysical Journal International* 145: 528534.

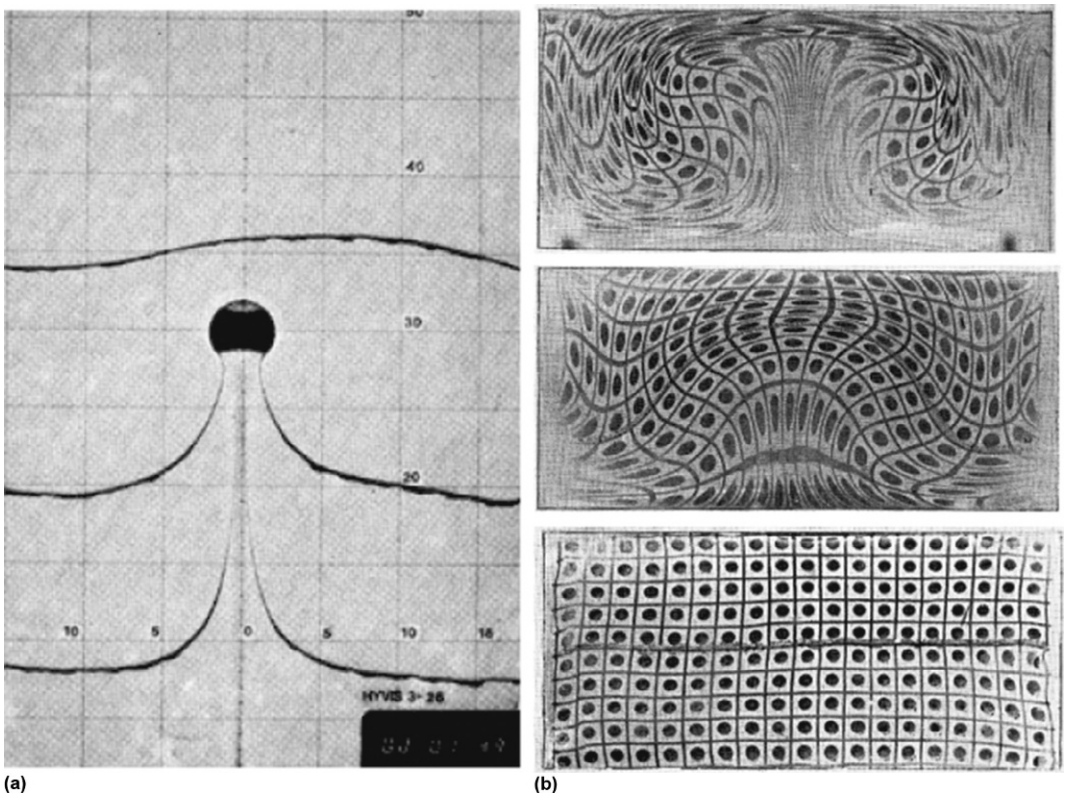


Figure 8 Fluid initially marked by horizontal dye streaks. (a) Rise of a thermal (reproduced from Griffiths RW (1986) Thermals in extremely viscous fluids, including the effects of temperature-dependent viscosity. *Journal of Fluid Mechanics* 166: 115138; Griffiths RW (1986) The differing effects of compositional and thermal buoyancy on the evolution of mantle diapirs. *Physics of the Earth and Planetary Interiors* 43: 261273). (b) Rise of a thermal plume. The gravity has been accelerated 2000 times in a centrifuge. Time increases from bottom to top pictures (reproduced from Weijermaars, 1988).

optical wave front, one can therefore deduce information on the refractive index, hence on the temperature or compositional field. The optical methods are divided into two groups: the shadow and schlieren techniques using the deflection of light in the measurement media (e.g., Settles, 2001) and interference methods based on differences of length of the optical paths, that is, on the phase (e.g., Hauf and Grigull, 1970). In the following, we will be referencing to a schlieren object, when we are considering the influence of this object on light deflection, and to a phase object when considering its influence on the optical paths length.

Both shadow and schlieren techniques are whole-field integrated optical systems that project line-of-sight information onto a viewing screen or camera focal plane. They use light intensity and light ray identification, so that only a good white light source is required. They are more appropriate for two-dimensional (2-D) or axisymmetric phenomena since they integrate the information along the light ray path but are still qualitatively useful for any phenomenon. Both techniques have a long history: R. Hooke developed a schlieren method as early as 1672 and the first published shadowgram of a turbulent plume over a flame by Marat dates back to 1783. Settles (2001) gave a comprehensive review of these two techniques.

Figure 9 Thermal boundary layer (TBL) instabilities in water visualized by an electrochemical technique. Adapted from Sparrow EM, Husar RB, and Goldstein RJ (1970) Observations and other characteristics of thermals. *Journal of Fluid Mechanics* 41: 793800.

7.03.3.1.2.1 Shadowgraph

Without the object present in the field of view, the light source illuminates the screen uniformly. With the object present, some rays are refracted, bent, and deflected from their original paths, producing a shadow. The optical inhomogeneities of the object redistribute the screen illuminance. The illuminance level responds to the second spatial derivative or Laplacian of the refractive index. Best results are obtained when the cell is illuminated by a (quasi-) parallel source of light (**Figures 5(a)** and **10(a)**). Due to the negative slope of the temperature dependence of the refractive indices of most liquids, the hot material acts as a diverging lens and appears dark (**Figure 10(b)**), while the cold material acts as a focusing lens and appears light (**Figure 10(c)**). Shadowgraph techniques are easy (and cheap!) to implement and allow the visualization of large objects. Since they image strong temperature gradients well, they have been used to determine convective patterns and measure their characteristic wavelength (e.g., Chen and Whitehead, 1968; Heutmaker and Gollub, 1987; see **Section 7.03.5**), measure thermal boundary layer (TBL) thicknesses (Olson et al., 1988), or follow the evolution of laminar thermal plumes (e.g., Moses et al., 1993;

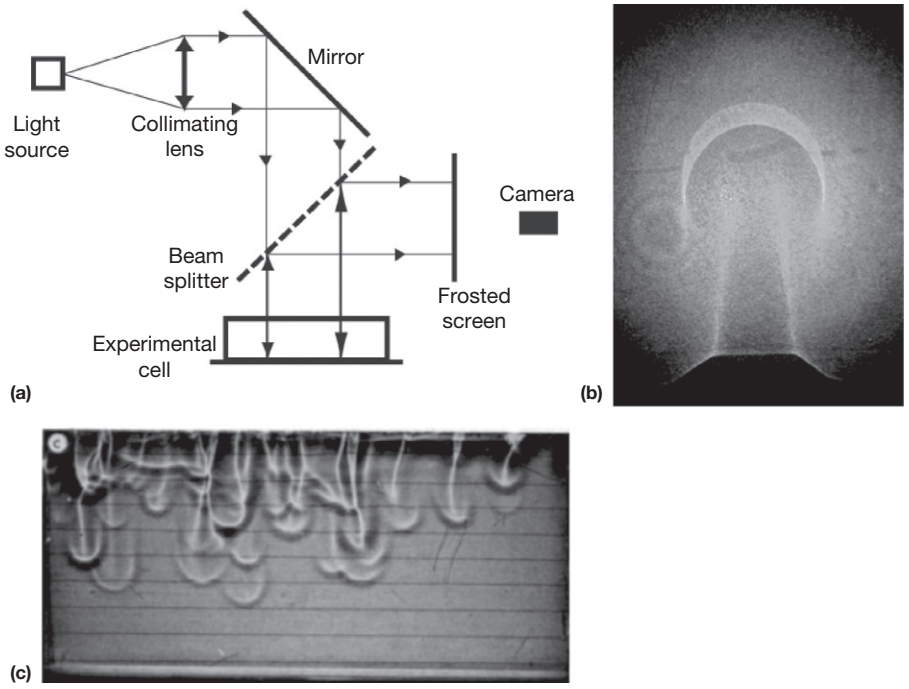


Figure 10 Shadowgraph. (a) Alternative setup to **Figure 5(a)** when the bottom plate of the cell is a mirror. (b) Hot thermal plume out of a small heat source. (c) Cold thermal instabilities from a cold horizontal plate (reproduced from Jaupart C, Brandeis G, and Allègre CJ (1984) Stagnant layers at the bottom of convecting magma chambers. *Nature (London)* 308: 535–538). Cold material focuses light and appear bright, while hot material acts as a divergent lens and appears dark.

Shlien, 1976). However, the whole temperature field cannot be retrieved from the shadowgraph images.

7.03.3.1.2.2 Schlieren

The schlieren image is a conjugated optical image of the schlieren object formed by a lens or a mirror. The method requires a knife-edge or some other sharp cutoff of the refracted light. **Figure 11(a)** presents the diagram of a simple schlieren system with a knife-edge placed at the focus of the second lens. Adding an object will bend light rays away from their original paths. The refracted rays miss the focus of the optical system. The upward-deflected ray brightens a point on the screen, but the downward-deflected ray hits the knife-edge causing a dark point against a bright background. So a vertical gradient of the refractive index is converted into an amplitude difference. In general, the illuminance level of the schlieren image responds to the first derivative of the refractive index on a direction perpendicular to the knife-edge. **Figure 11(b)** and **11(c)** shows the schlieren image of a hot plume obtained with a horizontal and a vertical knife-edge, respectively. **Figure 11(d)** shows the schlieren image of the same plume using a circular spatial filter of 0.5 mm diameter. The bright line in the center of the image and contours represent the zones with zero refractive index gradient.

For weak disturbances, the schlieren technique has a much higher sensitivity than the shadowgraph. It has been used to visualize the onset of thermal convection above a heated horizontal plate (Schmidt and Milverton, 1935), 2-D boundary layers along vertical walls (e.g., Hauf and Grigull, 1970), and shock waves (e.g., Settles, 2001). Another advantage over shadowgraph is the 1:1 image correspondence with the object of study. Moreover, 3-D study of the phase object is also possible by using a multisource system (Hanenkamp and Merzkirch, 2005).

Other optical methods can be derived from schlieren techniques. In the so-called lens-and-grid techniques, an array of light/dark stripes is used as a background grid. This simple method, also called background grid distortion, can be used when a large field of view is necessary with no need for high sensitivity. The sensitivity can be improved by adding another grid between the focusing lens and the image plane (**Figure 12**). The Moiré method can be seen as a variant of lens-and-grid method with grids on either sides of the schlieren object. Dalziel et al. (2000) contributed to this method by simulating electronically the second grid in the image capture device. Another variant called Moiré shearing interferometry (or Talbot interferometer) was used for mapping phase objects such as candle flames (Lohmann and Silva, 1972).

7.03.3.2 Temperature and Heat Flow Measurements

Since the precise characterization of thermal convection requires the quantitative knowledge of the temperature field, temperature measurements are a major step in data acquisition. Local measurements have long been provided by thermocouples or thermistors embedded within the flow, but with the risk of disturbing it. The last 30 years has seen the development of noninvasive methods. Laser beam scan and interferometry techniques can provide accurate temperature structure of 2-D or axisymmetric features. More recently, the use of thermochromic liquid crystals has allowed for the first time the accurate determination of a fully 3-D, time-dependent temperature field.

7.03.3.2.1 Local temperature measurements

Point temperature measurements can be obtained electronically using the variation with temperature of the electrical

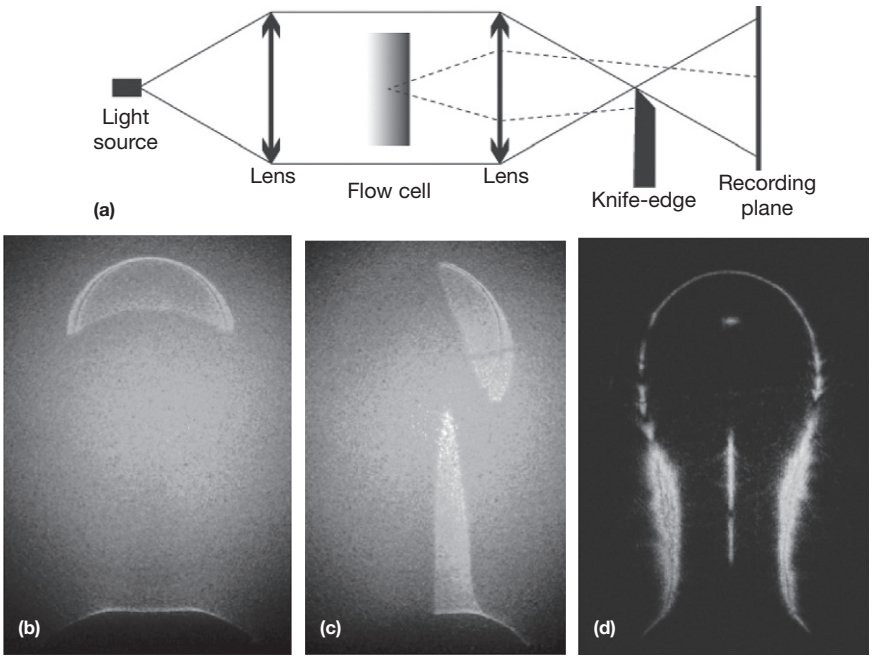


Figure 11 Schlieren. (a) Optical setup. (b)–(d) Pictures of a thermal plume out of a small heat source obtained with (b) a horizontal knife-edge, (c) a vertical knife-edge, and (d) a 0.5 mm diameter pinhole.

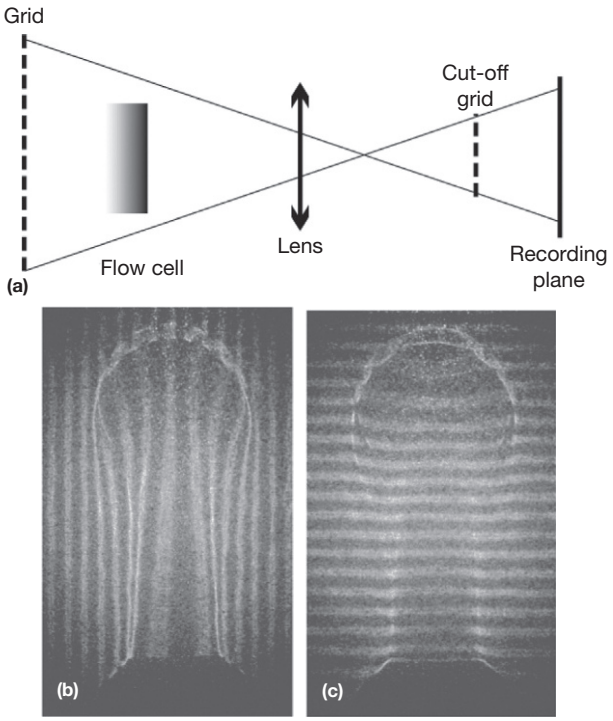


Figure 12 Lens-and-grid method. (a) Optical setup. (b and c) Same thermal plume as **Figure 9(b)** obtained with (b) a vertical grid and (c) a horizontal grid.

resistance of metals (Pt resistance thermometers and thermistors) or using the thermoelectric effect at the junction between two metals (thermocouples).

Pt thermometers exploit the increase with temperature of the electrical resistance of platinum. The most widely used sensor is the 100 or 1000 Ω platinum resistance thermometer. They are the most accurate and stable sensors over a long time period. However, they are expensive and do not come in diameters smaller than a few millimeters. So they are usually used to calibrate all the other temperature sensors in a lab, but are not used directly in the experimental cell.

Thermistors are made from certain metal oxides whose resistance decreases with increasing temperature. Their behavior is highly nonlinear, which limits their useful temperature span, and they are less stable than Pt thermometers. However, they are cheaper, and they can be produced in very small designs (0.1 mm) with a fast response and low thermal mass. The measurement of temperature then requires an electric power supply and a voltmeter. Properly calibrated thermistors can give a precision of 0.002 °C over a 10 °C range. They are used in high Rayleigh number convection experiments in water and lower-viscosity fluids (e.g., Castaing et al., 1989; Chilla et al., 2004; Niemela et al., 2000).

Thermocouples are based on the thermoelectric effect: the junction between two different metals produces a voltage that increases with temperature. In order for this thermal voltage to produce a flow of current, the two metals must also be connected together at the other end so that a closed circuit is formed. With different temperatures at the junctions, the voltages generated are different and a current flows. A thermocouple can thus only measure temperature differences between the measurement junction and the reference junction. The latter is at a known temperature, which is nowadays usually that of a commercial electronic ice-point cell. The temperature range of thermocouples is bigger than that for thermistors and their behavior is more linear. However, the voltage produced by the thermoelectric effect is very small. For the commonly used type K and type E, it amounts, respectively, to 40 and 60 μV per degree Celsius, so that a six-digit voltmeter will allow temperature measurements with an accuracy of ±0.025 °C. Their response time is about 2 s, that is, much shorter than the typical timescale of instabilities in viscous fluids.

The temperature sensors can be introduced permanently in the experimental tank either one by one at different locations (e.g., Davaille and Jaupart, 1993; Jellinek and Lenardic, 2009; Sparrow et al., 1970; Weeraratne and Manga, 1998) or on given vertical or horizontal profiles (e.g., Guillou and Jaupart, 1995; Le Bars and Davaille, 2002, 2004a,b). The latter allows one to follow the temperature profiles through time, but the number of sensors in a 2 mm diameter probe is limited to 14. Alternatively, the sensors can be mounted on a stepping motor and moved vertically to measure the vertical temperature profile (**Figure 5(b)**). With viscous fluids, a large volume of fluid is carried along each time the probe is moved to a new depth. Therefore, the probe must be kept several minutes at the same height in order for the system to equilibrate, and the vertical temperature profiles are determined from the time-averaged measurements at each depth (e.g., Giannandrea and Christensen, 1993; Matsumoto et al., 2006; Olson, 1984).

It is also possible to measure directly a horizontal average of the temperature field by stretching a set of very thin platinum wires (0.2 mm diameter) horizontally across the experimental cell (**Figure 10(c)**; Davaille and Jaupart, 1993; Giannandrea and Christensen, 1993; Jaupart and Brandeis, 1986). Because the wire resistance varies as a function of temperature, these wires are operated in the same way as thermistors, within a circuit made of a stable precision tension generator and a precision resistance. This setup is very delicate and time-consuming to calibrate but allows one to measure every second the horizontally averaged temperature with a 0.1 °C accuracy with a six-digit voltmeter. This method was used to study penetrative convection in constant-viscosity fluids (Jaupart and Brandeis, 1986) and in strongly temperature-dependent viscosity fluids (Davaille and Jaupart, 1993, 1994).

The advantages of local temperature sensors are their high precision and the ability to obtain long time series (high-frequency sampling and/or long run) at low cost. However, the sensors are introduced within the fluid, which imposes a zero-velocity condition for the flow on their surface. They therefore always perturb the flow locally. For nonsteady flow and thin probes, this perturbation remains negligible and the probes do not preferentially focus convective features on themselves (e.g., Davaille and Jaupart, 1993). However, for steady flow (like the low Rayleigh number convection), upwellings or downwellings can be locked on the temperature probes. In such cases, one should use only removable temperature sensors mounted on stepping motors and/or the nonintrusive methods, which we shall describe in **Sections 7.03.3.2.2–7.03.3.2.4**.

Taking advantage of the advances in miniature sensors and communication devices, Lagrangian 'smart particles' have

recently been developed (Shew et al., 2007). It consists of a small (2 cm diameter) neutrally buoyant capsule equipped with onboard electronics, which measures temperature and transmits the data via a wireless radio frequency link to a desktop computer. The device resolves millikelvin changes in temperature with up to 100 ms sampling time. The temperature sensor is carried along the flow and its position can be tracked using a camera. This technique has been used to obtain statistics of velocity, temperature, and heat transport dynamics in turbulent Rayleigh–Bénard convection (Gasteuil et al., 2007). However, given the size of the particle, determination of the TBL characteristics remains problematic.

7.03.3.2.2 Deflection of a light beam: 1-D profile

In the case of 2-D or axisymmetric refractive index distributions, it is possible to recover the temperature field by scanning the test zone with a laser beam and recording its deflection. Nataf et al. (1988) and Rasenat et al. (1989) applied this method to determine the type of coupling in a two-layer convection at low Rayleigh numbers. Laudenbach and Christensen (2001) described the method in detail for the axisymmetric case (**Figure 14**) and applied it to thermal plume conduit and solitary waves.

The same treatment can be applied to the lens-and-grid methods described in **Section 7.03.3.1.2**, since the ray displacement can by calculated from the grid distortion provided that the structure is 2-D or axisymmetric. The spatial resolution of the method then depends on the spatial frequency of the grid lines.

7.03.3.2.3 Interferometry: 2-D and 3-D

Compared to shadowgraph and schlieren methods, interferometric methods offer more detailed information about the phase object (e.g., Yang, 1989), since they play with the light path, but they usually require laser light and a good optical control. They have been used for pattern determination (e.g., Jaupart et al., 1984; Kaminski and Jaupart, 2003; Nataf et al., 1981, 1984, 1988) and for quantitative measurement of the temperature field (e.g., Gebhart et al., 1970; Shlien and Boxman, 1981).

7.03.3.2.3.1 Mach–Zehnder interferometry

The most common two-beam interferometer is the Mach–Zehnder interferometer (**Figure 13(a)**; Hauf and Grigull, 1970). The phase object is placed in one of the legs of the interferometer formed by two-beam splitter mirrors and two total reflecting mirrors. For a 2-D object, the interference fringes can be interpreted as isothermal lines. It has been used for boundary layer dynamics along vertical walls or in double-diffusive systems (e.g., Lewis et al., 1982) and steady laminar plumes above a horizontal line heat source (**Figure 13(b)**; Gebhart et al., 1970). For a 3-D object, the phase difference between the two optical paths is integrated over the longitudinal length of the object, and the interferogram is not readily interpretable (Merzkirch, 1993). In a cylindrical symmetry, the radial temperature distribution can be obtained rather simply by an Abel transformation (Vest, 1975). If the refractive index has no symmetry, tomography algorithms and reconstruction techniques have to be applied. The Mach–Zehnder interferometer was used to measure the temperature field of axisymmetric, laminar thermal plumes in liquids (**Figure 13(c)**; e.g., Boxman and Shlien, 1978; Chay and Shlien, 1986; Shlien and Boxman, 1979, 1981) and for turbulent mixing of salt solutions (Boxman and Shlien, 1981). More recently, Qi et al. (2006) used the Michelson interferometer, more often used to measure surface displacement (see **Section 7.03.3.5**), to measure the temperature field of 3-D axisymmetric flames.

Holographic interferometry was introduced in flow visualization by Heflinger et al. (1966) and Tanner (1966). Holography has opened a new dimension for two-beam interferometry: the reference and test beams can be separated in time rather than space. In a holographic interferometer, two consecutive exposures are taken through the field of interest, usually the first exposure without object and the second in the presence of an object of varying temperature. The double exposed plate is developed and placed in the holographic reference beam for reconstruction. The pattern that becomes visible after this reconstruction is equivalent to the one obtained by a Mach–Zehnder interferometer (Hauf and Grigull, 1970; Merzkirch, 1993). The

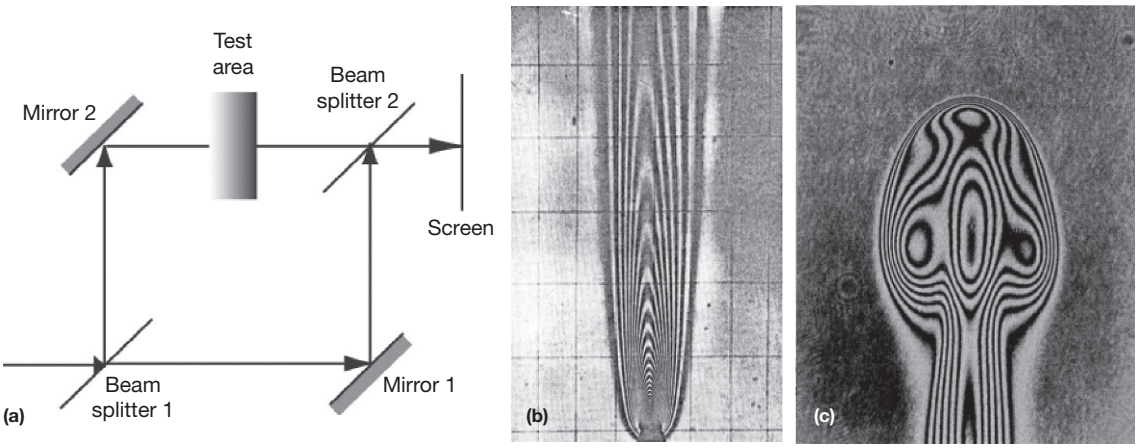


Figure 13 Mach–Zehnder interferometry. (a) Setup. (b) 2-D laminar plume from a heat line (reproduced from Gebhart B, Pera L, and Schorr AW (1970) Steady laminar natural convection plumes above a horizontal line heat source. *International Journal of Heat and Mass Transfer* 13: 161171). (c) Axisymmetric laminar plume from a point source in water (reproduced from Shlien DJ and Boxman RL (1981) Laminar starting plume temperature field measurement. *International Journal of Heat and Mass Transfer* 24: 919–930).

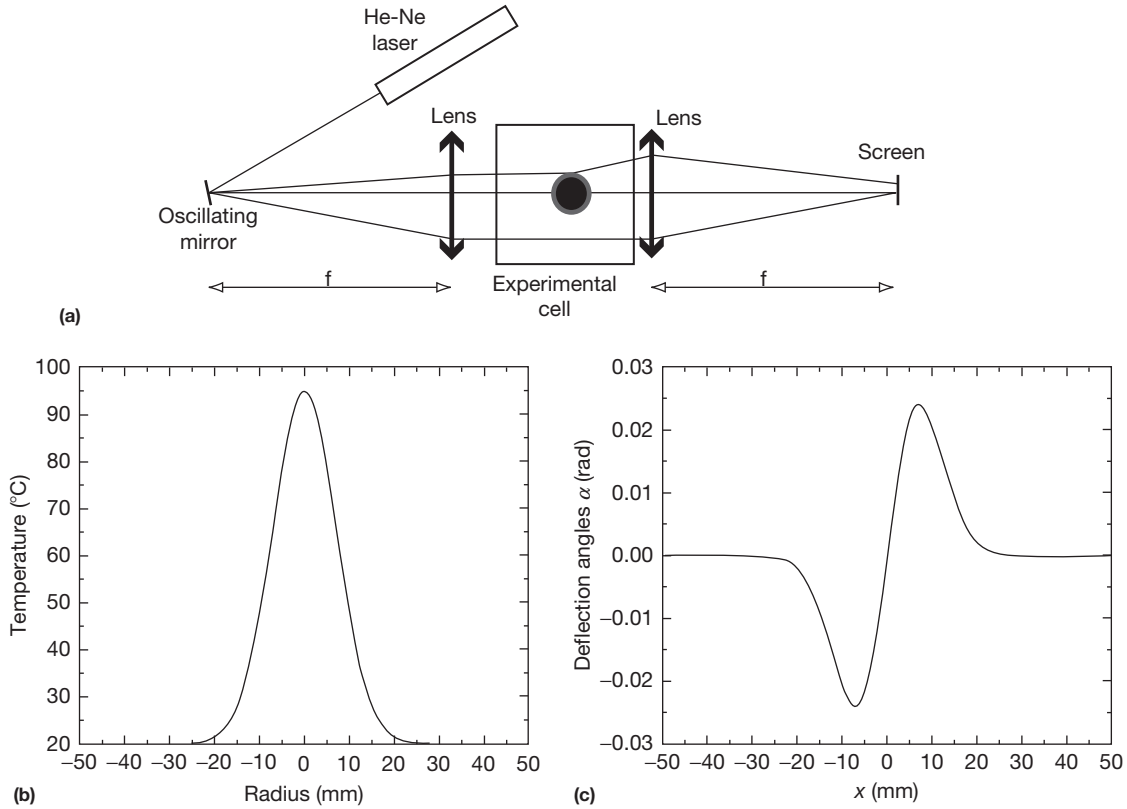


Figure 14 Deflection of a laser beam by a thermal plume. (a) Experimental setup. (b) Radial temperature structure of the plume. (c) Beam deflection angle. Adapted from Laudenbach N and Christensen UR (2001) An optical method for measuring temperature in laboratory models of mantle plumes. *Geophysical Journal International* 145: 528534.

main advantage of holographic interferometry over classical interferometers is that, since the geometric path of the signal in the two exposures is identical, the quality of the optical components is not crucial. The eventual optical disturbances and impurities in the test section cancel out.

7.03.3.2.3.2 Differential interferometry

Differential interferometry (also called shearing interferometry) is a method to measure derivatives of light phase distortions. In practice, this can be done in two ways. One possibility is to send two beams slightly displaced through a phase object and to put them back together on the camera. The best way for doing this is to use two Wollaston prisms and polarizers (Oertel and Buhler, 1978). Another possibility is to divide one wave into two identical beams behind the object and to put them on a camera with a small displacement. This can be done, for instance, by using one Wollaston prism (Sernas and Fletcher, 1970; Small et al., 1972), as in the GeoFlow experiment (**Figure 6**). Another variant is the use of a Mach–Zehnder interferometer with the phase under investigation placed outside (Pretzel et al., 1993). In this arrangement, using a very compact Mach–Zehnder interferometer minimized its inherent sensitivity to vibrations. Another simple way of separating into two beams is to use an optical flat plate tilted at a certain angle (**Figure 15**; Davaille et al., 2011; Jaupart et al., 1984; Kaminski and Jaupart, 2003). Extinction lines follow constant horizontal temperature gradients. Vertical extinction lines away from the plume correspond to light beams through uniform background (**Figure 15**).

The main advantage of differential interferometry is its variable sensitivity: carrier fringe orientation and frequency can be chosen separately. It is therefore possible to visualize with the same apparatus the flow configuration in two fluids with very different temperature dependence of the refractive index (e.g., Oertel and Buhler, 1978). Moreover, the separation between the object and the interferometer facilitates the investigation of large and complicated objects. The evaluation of the interferogram can be done by Fourier analysis, and the result, being the first derivative of the integral phase shift caused by the object, is obtained with high accuracy. For radially symmetrical objects, the spatial distribution of the refraction index can be obtained easily because the Abel inversion formula is reduced to a simple integration, which can be done more reliably than a differentiation. In that respect, differential interferometry results are less noisy than Mach–Zehnder interferometry, where the spatial distribution of the refractive index is obtained through numerical differentiation of the optical path difference. More details on the digital processing of interferograms can be found in the monograph by Yang (1989).

Speckle interferometry can also enter into this category (Merzkirch, 1995). In double-exposure speckle photography, two speckle patterns are recorded on the same photographic plate. Between the two exposures, the scattering plate (ground glass) is shifted to obtain the interferometric fringes. After

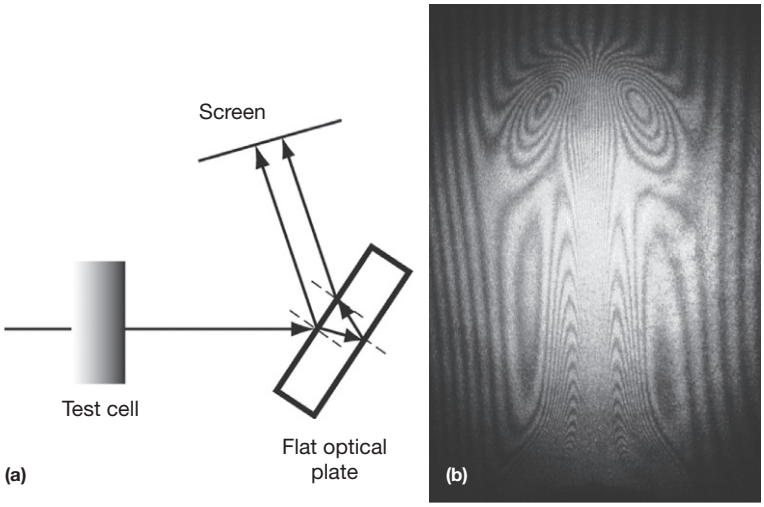


Figure 15 Differential interferometry. (a) Setup. (b) Thermal plume in sugar syrup.

photographic development, the specklegram is scanned by a laser beam. By measuring Young's fringe spacing and orientation, it is possible to measure the two components of the displacement and convert them into deflection angles (resembling for this reason the shadowgraph and schlieren methods). In addition, precise multiprojection speckle photography allows the reconstruction of a 3-D temperature field using computer tomography (Asseban et al., 2000). Among the most important advantages of speckle photography are its spatial resolution (about 0.2 mm) and the possibility to collect a great amount of experimental information from a single specklegram. Large amounts of data can be processed and analyzed statistically. The potential of analyzing spatial characteristics of turbulent flows was demonstrated (Merzkirch et al., 1998).

7.03.3.2.4 3-D fields: thermochromic liquid crystals and fluorescence

7.03.3.2.4.1 Thermochromic liquid crystals

The use of thermochromic liquid crystals allows one to visualize the temperature field on a 2-D plane in the fluid flow without perturbing it (Dabiri and Gharib, 1991; Rhee et al., 1984; Willert and Gharib, 1991). Liquid crystals are mesomorphic phases that present peculiar properties due to the presence of some degree of anisotropy (Chandrasekhar, 1977). One particular class of these mesophases, chiral nematics (cholesterics), has a structure that undergoes a helical distortion, and because of their periodic structure, they generate Bragg reflections at optical wavelengths. The pitch of the wavelength of the Bragg-reflected light depends on the temperature T (de Gennes and Prost, 1993). Thus, the color of the material can change drastically over a temperature interval of a few degrees. When the liquid crystals are illuminated by white light, their color changes with increasing temperature from colorless to red at low temperatures, passes through green and blue to violet, and turns colorless again at high temperatures. It was first used as paint on a surface to determine convective patterns qualitatively (e.g., Chen and Chen, 1989; Lithgow-Bertelloni et al., 2001).

Subsequently, it became possible to mix the thermo liquid crystal (TLC) slurry directly within the fluid and to illuminate the tank cell on cross sections (**Figure 16(a)**). One of the main applications of this technique has been the study of aqueous turbulent flows (Dabiri and Gharib, 1996; Gluckman et al., 1993; Moses et al., 1993; Park et al., 2000; Pottebaum and Gharib, 2004; Solomon and Gollub, 1990, 1991; Zhou et al., 2007). The use of this method to measure the temperature field quantitatively requires a high-precision color CCD camera and a very precise calibration of the color of the liquid crystal particles against the true temperature. Moreover, the total temperature range accessible with one particular TLC slurry is usually around 2–3 °C. This technique has lately been further extended to the joint measurement of temperature and velocity in a 3-D field (Ciofalo et al., 2003; Fujisawa and Funatani, 2000; Kimura et al., 1998). An uncertainty analysis performed by Fujisawa and Hashizume (2001) gives an error <0.1 °C, or 5% of the total imposed temperature difference, for a calibration method based on a hue–saturation–intensity approach (Fujisawa and Adrian, 1999).

Convection in conditions analogous to those of the mantle involves viscous fluids (see **Section 7.03.2**) and typical temperature heterogeneities ~10–25 °C. Therefore, the use of one TLC slurry is not enough to image the whole temperature field. White light and a single TLC were used to determine the influence of internal heating (Takahashi et al., 2010; Tasaka et al., 2005), mechanical (Namiki and Kurita, 1999, 2001), and thermal boundary heterogeneities (Matsumoto et al., 2006) on thermal convection. The type of TLC was chosen to image well the hot instabilities. This permitted quantification of the convective pattern, but not accurate measurement of the temperature field. Illuminating several TLC slurries by a single laser wavelength plane sheet and recording the images with a high-precision black and white CCD camera (**Figure 16(a)**), Davaille et al. (2008, 2011) demonstrated that it was possible to obtain isotherms and local temperature gradients. The laser beam was expanded into a beam using a cylindrical lens. Each slurry brightens over a different temperature subrange and

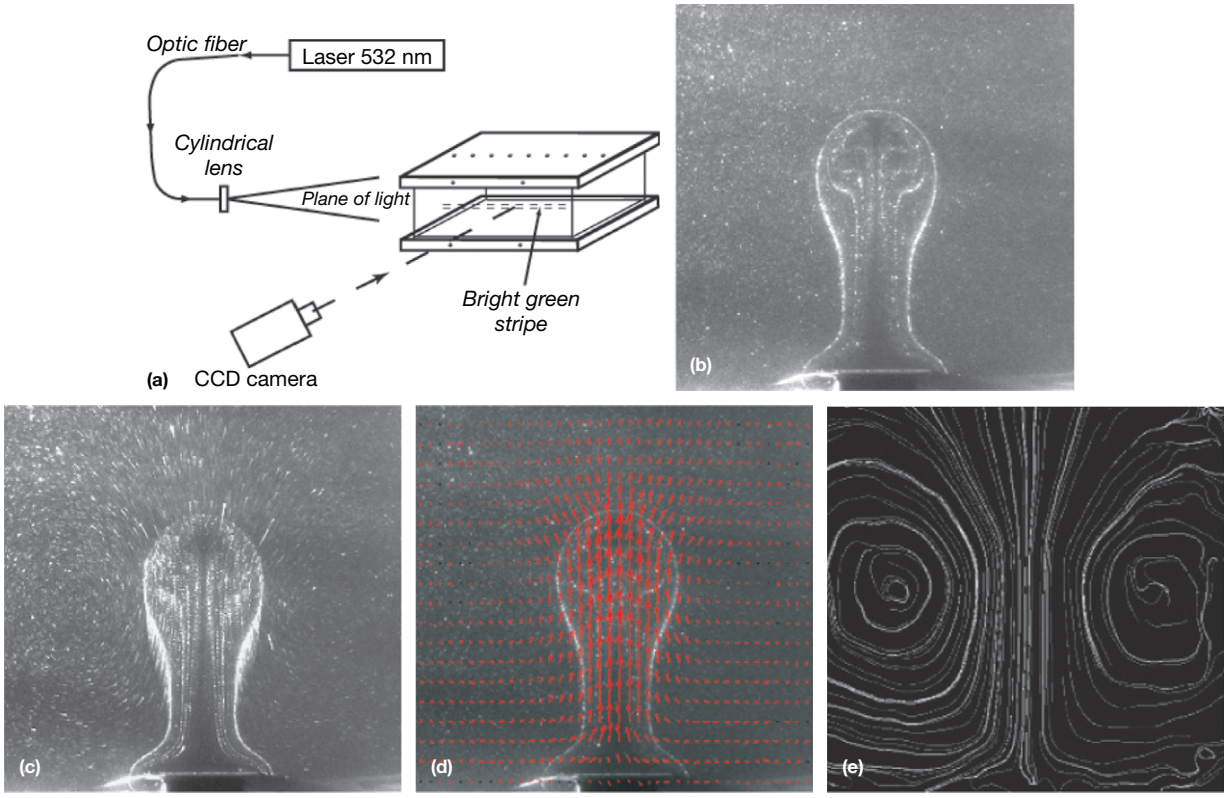


Figure 16 Temperature and velocity field of a thermal plume out of a localized heat source. (a) Setup, (b) isotherms, (c) streaklines taken over 10 s, (d) velocity field calculated by PIV, and (e) streamlines calculated from the velocity field.

therefore generates a horizontal bright line (**Figure 16(b)**). Each stripe presents a finite thickness: although each TLC responds at a given wavelength to a precise temperature, the polymeric capsules enclosing them introduce a scatter around this value. On a plot of the intensity as a function of depth, or temperature, each stripe corresponds therefore to a peak whose maximum defines the value of the 'isotherm' and whose thickness gives a measure of the local temperature gradient. After calibration, the light intensity maximum gives the 'isotherm' temperature with a precision of $\pm 0.1\,^{\circ}C$, and the half width of the bright lines gives the local temperature gradient. This method was used to study patterns of convection (Androvandi et al., 2011; Kumagai et al., 2007, 2008; Le Bars and Davaille, 2002, 2004a,b), plumes arising from a viscous TBL (Davaille and Vatteville, 2005), and plumes from a point heat source (Davaille et al., 2011, 2013; Massmeyer et al., 2013; Vatteville, 2004; Vatteville et al., 2009). Since viscous fluid motions are slow, it is possible to scan the experimental cell with the laser sheet (by means of an oscillating mirror driven by a galvanometer) to obtain the 3-D structure of the temperature field. This method can be combined with laser-induced fluorescence (LIF) and particle image velocimetry (PIV) to obtain simultaneously the composition (see **Section 7.03.3.3**) and velocity fields (see **Section 7.03.3.4**).

7.03.3.2.4.2 Two-color LIF

The fluid can also be doped with a low concentration of a fluorescent organic dye, like rhodamine B, and illuminated with a 532 nm laser sheet. The rhodamine B fluorescence intensity depends on temperature, concentration, and the spectral band of observation. Then, the ratio collected on two different spectral bands of detection, which have different temperature sensitivities, enables the removal of the unknown effects of laser intensity, fluorescent tracer concentration, and probe volume dimensions (e.g., Lavieille et al., 2001). The same type of measurements can be performed using thermographic phosphor materials and illumination with ultraviolet light (Abram et al., 2013).

7.03.3.2.5 Heat flow measurements

The dependence of the global heat transport on convection characteristics is one of the fundamental questions of Rayleigh–Bénard studies (see **Section 7.03.5**).

The most accurate measurement of the global heat flow is obtained through the measurement of the electric power needed to heat the bottom plate (e.g., Brown et al., 2005; Funfschilling et al., 2005; Giannandrea and Christensen, 1993; Malkus, 1954; Schmidt and Milverton, 1935; Silveston, 1958). With careful design of the experimental tank in order to minimize the heat losses (cf. **Section 7.03.2.3**), late results have reached a precision of 0.1% (e.g., Funfschilling et al., 2005).

The heat flow can also be deduced from the first derivative of the measured temperature profiles, either locally (using temperature probes or isotherms) or horizontally averaged (using platinum wires). This method requires a precise

knowledge of thermal conductivity (especially as a function of temperature) and temperature measurements at several different heights within the TBLs close to the horizontal plates to resolve the steep gradients there. This becomes increasingly difficult as the Rayleigh number increases and the boundary layers become thinner, with the danger of underestimating the heat flux. For Rayleigh numbers up to 10^8, a precision of 5% has been obtained (e.g., Davaille and Jaupart, 1993).

7.03.3.3 Composition

Measurement of the time evolution of composition is required in studies on entrainment, mixing, and two-layer convection. It usually amounts to measuring a tracer dilution. Point-based techniques use either *in situ* probes (e.g., conductivity) or extraction of samples by suction from various points in the flow field. Tait and Jaupart (1989) used conductivity measurements to study the mushy crystallization of ammonium chloride in viscous solutions. Davaille (1999a,b) studied entrainment processes in two-layer convection by periodic sampling of the fluid layers, measuring the salinity of each sample using the dependence of the refractive index on the salt concentration, and a food dye concentration by a UV absorption technique. However, these point-based techniques have some major drawbacks: the flow may be disturbed by the probes, the number of measurement points is very limited, and extraction techniques can yield only time-averaged concentrations and cannot capture their instantaneous fluctuations.

Concentration profiles averaged over the experimental cell (e.g., Olson, 1984) or over one of its dimension (e.g., Solomon and Gollub, 1988a,b, 1991) can be deduced from the optical absorption by the cell of an extended white light beam (e.g., Olson, 1984) or a laser beam (e.g., Solomon and Gollub, 1988a,b, 1991).

The advent of LIF techniques in the 1970s enabled simultaneous capture of the entire instantaneous tracer (fluorescent dye) concentration field over a planar sampling area (laser sheet), with an experimental setup similar to **Figure 16(a)**. LIF is a nonintrusive technique, which has been applied to turbulent jet flows (e.g., Villermaux and Innocenti, 1999), laminar mixing (e.g., Fountain et al., 2000; Ottino et al., 1988; Seon et al., 2006), and two-layer convection (e.g., Kumagai et al., 2007, 2008). Since fluorescein dye shifts the laser light frequency, it is possible to use LIF and TLCs to measure composition and temperature simultaneously, by recording the images with two different filters, one for the laser frequency and one for the fluorescein (**Figure 17**; Kumagai et al., 2007). It is also possible to determine simultaneously temperature and concentration from LIF alone. To deconvolute the two signals, three different spectral bands of observation are used (Maqua et al., 2006).

7.03.3.4 Fluid Motions

7.03.3.4.1 Local measurements

It is easy to follow the displacement of an interface defined by a grid (e.g., Weijermars, 1988) or by abrupt gradients either in refraction index (e.g., visualized by shadowgraph or by interferometry) or in dye concentration. These techniques have been used to measure the rising velocity of thermals

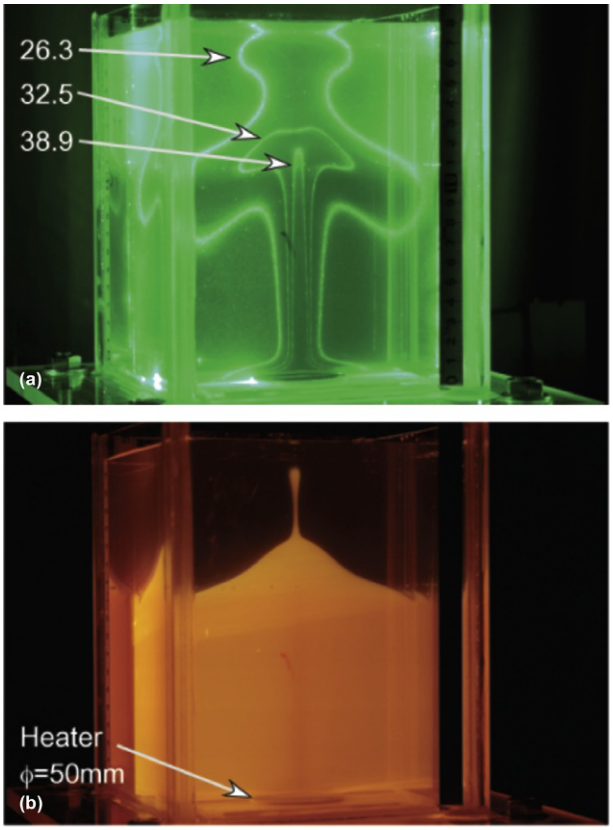


Figure 17 Interaction of a thermal plume with a density stratification. The denser layer has been dyed with fluorescein. 2-D visualization when the tank is illuminated with a laser sheet (532 nm): (a) isotherms (26.3, 32.5, and 38.9 °C) and (b) compositional fields (laser-induced fluorescence, LIF method). These two images were taken by a three-CCD camera at almost same time, but using two different optical filters, a band-pass filter for temperature and a cutoff filter for composition. Reproduced from Kumagai I, Davaille A, and Kurita K (2007) On the fate of mantle thermal plumes at density interface. *Earth and Planetary Science Letters* 254: 180–193.

(Griffiths, 1986a,b) or plumes (e.g., Coulliette and Loper, 1995; Griffiths and Campbell, 1990; Kaminski and Jaupart, 2003; Moses et al., 1993; Olson and Singer, 1985).

Hot wires probes can also be used whereby the local flow velocity is measured by sensing the rate of cooling of fine, electrically heated wires. However, their use is usually confined to turbulent or high-speed flows of low-viscosity fluids. Moreover, as an intrusive method, it suffers the same drawbacks as temperature sensors (cf. **Section 7.03.3.2.1**).

Laser Doppler velocimetry (LDV) allows one to measure continuously in time and at a given position in space up to the three components of the velocity of tracer particles (e.g., Adrian, 1983; Merzkirch, 2000). The method is based on the optical Doppler effect and requires seeding of the working fluid with μm size particles. Incident light is scattered by the moving particles, and the frequency of the scattered radiation is Doppler-shifted. Since this frequency shift is relatively small, its detection requires the use of monochromatic incident laser light. Maps of the spatial dependence of the velocity field can be obtained by translating the cell apparatus over

the LDV system. Solomon and Gollub (1988a,b, 1991) applied this technique to measure the velocity field in the steady quasi-2-D Rayleigh–Bénard convection.

7.03.3.4.2 2-D and 3-D field measurements

One of the oldest techniques for measuring velocities in a fluid is seeding particles (e.g., hollow glass beads, aluminum flakes, and tiny air bubbles) into the fluid and illuminating a cross section of the flow cell by a thin sheet of light. The foreign particles should be small and have densities as close to that of the fluid as possible, in order to follow passively the local flow. There are then several ways to describe the flow. The trajectory of a single fluid particle over time defines a pathline. A streakline connects all the fluid elements that have passed through a given point (**Figure 16(c)**). Streamlines are tangential to the flow directions at a given time (**Figure 16(e)**). In steady flow, pathlines, streaklines, and streamlines are identical, but usually not in time-dependent flow.

The availability of high-power laser sources together with fast digital processors led to the development of sophisticated whole-field velocimetry techniques such as particle image velocimetry (PIV) and particle tracking velocimetry (PTV). Both techniques provide a quantification of the velocity field over the entire plane. The optical setup is the same as for the visualization of isotherms (**Figure 16(a)**). PTV tracks individual particles in subsequent images, while PIV follows a group of particles through statistical correlation of local windows of the image field from two sequential images (Adrian, 1991; Merzkirch, 2000; Raffel et al., 1998). From the known time difference and the measured displacement, the velocity is calculated. PTV is convenient when the seeding is sparse, and/or near boundaries, since each particle trace is analyzed individually. The latter operation is time-consuming. Moreover, as the particles are present at random locations in the fluid, the velocity estimators are found at random locations in the flow, too.

The PIV scheme, on the contrary, requires a high particle density and calculates the mean displacement of particles in a small region of the image (the interrogation window) by cross correlation of the transparency signal of the same window in two subsequent images. It therefore removes the problem of identifying individual particles, which is often associated with tedious operations and large errors in the detection of particle pairs. Its spatial resolution is uniform (function of the interrogation window size) over the whole image. In geodynamics, this technique has recently been used to study subduction (Kincaid and Griffiths, 2003, 2004; Schellart, 2008) and thermal convection (Androvandi et al., 2011; Davaille and Vatteville, 2005; Kumagai et al., 2007, 2008). Current research aims at combining the resolution of PTV with the fast algorithm of PIV. Funiciello et al. (2006) recently developed such a feature tracking method to map mantle flow during retreating subduction.

PIV has also been used simultaneously with TLCs in two ways. When the temperature difference is small and only one TLC slurry is needed to determine a continuous temperature field, the velocity is measured using the same TLC particles as Lagrangian flow tracers (e.g., Ciofalo et al., 2003; Dabiri and Gharib, 1991; Park et al., 2000). When several TLC slurries are used to visualize isotherms, the fluid can also be seeded uniformly with 10 μm size particles (hollow glass or latex) to calculate the PIV (**Figure 16**; Davaille and Vatteville, 2005). 3-D determination of the two fields is possible by scanning the experimental tank with the laser sheet (e.g., Ciofalo et al., 2003). The velocity field can also be obtained in a volume by stereoscopy, whereby two cameras record the particle displacements from different angles.

7.03.3.4.3 Stress and strain rate

Stress and strain rate can be determined from the displacement and velocity fields measured by the methods just described; indeed, this is the most common method.

However, one can also use optical streaming birefringence. This effect, already known to Mach (1873) and Maxwell (1873), is the property of certain liquids to become birefringent under the action of shear forces in a flow. Such fluids consist mainly of elongated and deformable molecules (e.g., polymers) or contain elongated, solid, crystal-like particles in solution. At rest, these particles are randomly distributed and the fluid is optically isotropic. The shear forces in a flow cause the particles to align in a preferential direction, which renders the fluid anisotropic. This phenomenon is also well known in solids (and used in photoelasticity) and in the mantle is responsible for seismic anisotropy. Light propagation in such a medium is directionally dependent: an incident light wave separates in the birefringent liquid into two linearly polarized components whose planes of polarization are perpendicular to each other and which propagate with different phase velocities. Different values of the refraction index are therefore assigned to the two components. They are out of phase when leaving the birefringent liquid, and this difference in optical path can be measured by interferometry.

Although theoretically most fluids should show this effect, the best results have been obtained by using milling yellow dye dissolved in aqueous solutions (for its physicochemical properties, see Pindera and Krishnamurthy, 1978; Swanson and Green, 1969). Milling yellow solutions can be strongly non-Newtonian, depending on concentration and temperature. One special concentration is particularly interesting, since its flow curve is similar to that of human blood (Schmitz and Merzkirch, 1981).

Interferometric fringes, 'isochromates,' can then be obtained with a Mach–Zehnder interferometer or with a polariscope (**Figure 18**). However, a theoretically based flow-optic relation that would allow quantitative determination of the flow from the refraction index distribution does not yet exist. Empirical relations concern only small values of the maximum strain rate, or small Reynolds numbers (creeping flow). For 2-D or axisymmetric situations, a linear relationship between the refractive index differences and the strain rate has been successfully used to determined the flow in pipes of varying cross sections (e.g., Schmitz and Merzkirch, 1981). Horsmann and Merzkirch (1981) developed a flow-optic relation that applies to the general 3-D case. More details on the technique can be found in Merzkirch (1989).

7.03.3.5 Surface Topography

Mantle convection produces sizable topography on Earth. Hot spots are usually located on top of broad topographic swells (see Ballmer et al., **Chapter 7.10**). Experiments on the interaction between a rising buoyant plume and the lithosphere, commonly proposed to explain hot spot volcanism, were the

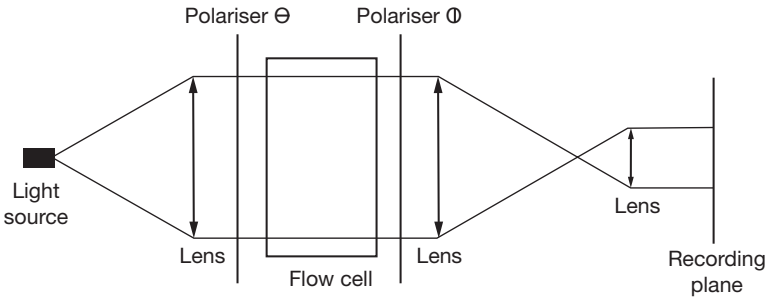


Figure 18 Streaming birefringence. Polariscope setup. Reproduced from Merzkirch W (1989) Streaming birefringence. In: Yang W-J (ed.) *Handbook of Flow Visualization*, pp. 177–180. London: Taylor & Francis.

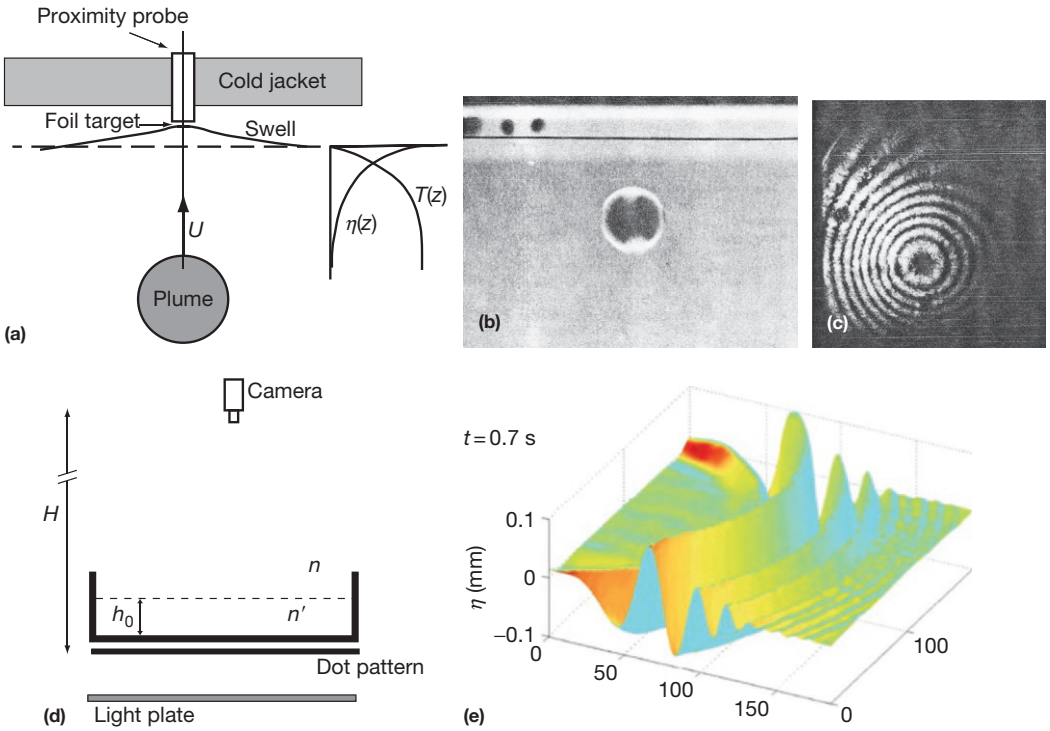


Figure 19 (a–c) Measurements of surface displacements above a rising drop. (a) Proximity probe (sketch adapted from Olson P and Nam IS (1986) Formation of seafloor swells by mantle plumes. *Journal of Geophysical Research* 91: 71817191). (b) Side view of the rising drop. (c) Holographic interferometry on the free surface. The center of the interference fringes corresponds to the surface highest point and is centered on the drop vertical axis (reproduced from Griffths R.W, Gurnis M, and Eitelberg G (1989) Holographic measurements of surface topography in laboratory models of mantle hotspots. *Geophysical Journal* 96: 1–19). (d and e) FS-SS method: (d) Sketch of the experimental setup ($H/h_0 \sim 1000$). (e) Perspective view of the circular waves created by the impact of a drop. Reproduced from Moisy F, Rabaud M, and Salsac K (2009) A synthetic schlieren method for the measurement of the topography of a liquid interface. *Experiments in Fluids* 46: 1021–1036.

first to involve measurements of the upper surface displacement. Two techniques were used. Olson and Nam (1986) measured the amplitude of the surface topography above the axis of a rising buoyant drop using a high-frequency induction coil proximity probe (**Figure 19(a)**). They were able to follow the temporal evolution of the surface elevation with a sensitivity of 0.02 mm. Using a Michelson interferometer, Griffths et al. (1989) could reconstruct from the fringe pattern both the shape and the width of the swell, in addition to its height (**Figure 19(b)**). Fringes in the interferogram are topographic contours with a vertical distance apart of half the laser wavelength (i.e., 0.633 μm for a He–Ne laser). The surface height at any position of the swell is therefore given by the number of fringes.

More recently, the use of commercial 3-D laser scanner provides topography with a precision of 0.1 mm (e.g., Bajolet et al., 2012), but one full image takes several minutes to scan, and this method is therefore reserved to slowly varying phenomena. When the phenomenon timescale is shorter, there are two other interesting methods, for which software packages now exists. The Moiré method is based on projecting a fringe

pattern on the surface and analyzing the deformation of the pattern caused by the topography with respect to a reference plane (Pouliquen and Forterre, 2002). Under optimized experimental conditions, topography uncertainties are between 20 and 100 μm with submillimeter spatial resolution (Limare et al., 2011). To measure the instantaneous deformations of the interface between two transparent fluids, the free-surface synthetic schlieren (FS-SS) method (**Figure 19(d)** and **19(e)**) is based on the analysis of the refracted image of a random dot pattern visualized through the interface. The apparent displacement field between the refracted image and a reference image obtained when the surface is flat is determined using a digital image correlation algorithm. A numerical integration of this displacement field is then used for the reconstruction of the instantaneous surface height. Deformations as low as 1 μm can be measured, with a vertical resolution ~2% of the maximum deformation, with a low computational cost (e.g., **Figure 19(e)**). However, this technique is unable to determine the displacement field for a strong curvature and/or large surface-pattern distance (Moisy et al., 2009).

7.03.4 Rayleigh–Taylor Instabilities

Whenever light fluid underlies a heavier fluid in a field of gravity, the interface between them is inherently unstable to small perturbations (**Figure 1(a)**). This is the classic form of the so-called Rayleigh–Taylor instability (hereafter 'RTI'), first studied theoretically by Rayleigh (1883) and later by Taylor (1950) (see Ribe, **Chapter 7.04**). RTIs have been used to model a number of geophysical processes, including the formation and distribution of salt domes (e.g., Biot and Ode, 1965; Nettleton, 1934; Ribe, 1998; Selig, 1965); the emplacement of gneissic domes and granitic batholiths (Fletcher, 1972); the instability of continental lithosphere beneath mountain belts (Houseman and Molnar, 1997); the subduction of oceanic lithosphere (Canright and Morris, 1993); the temporal and spatial periodicity of volcanic activity in a variety of geologic settings, namely, island arcs (Fedotov, 1975; Kerr and Lister, 1988; Marsh, 1979; Marsh and Carmichael, 1974), continental rifts (Bonatti, 1985; Mohr and Wood, 1976; Ramberg and Sjostrom, 1973), Iceland (Sigurdsson and Sparks, 1978), and mid-ocean ridges (Crane, 1985; Kerr and Lister, 1988; Schouten et al., 1985; Whitehead, 1986; Whitehead et al., 1984); the segregation and mixing in the early history of the Earth's core and mantle (Jellineck et al., 1999); the initiation of instabilities deep in the mantle (Bercovici and Kelly, 1997; Kelly and Bercovici, 1997; Loper and Eltayeb, 1986; Ramberg, 1972; Ribe and de Valpine, 1994; Stacey and Loper, 1983; Whitehead and Luther, 1975); the survival of a sheetlike mantle upwelling (Namiki et al., 2013); and the instabilities of a tilted plume conduit (Kerr and Lister, 2008; Kerr et al., 2008; Lister et al., 2011; Olson and Singer, 1985).

Laboratory experiments have proved to be powerful tools for studying the development and morphology of RTI. We focus here on the case where the density field is nondiffusing and the fluids are highly viscous, so that the effects of compositional diffusion, inertia, and surface tension can be neglected. Then, the nature of overturning only depends on the geometry of the boundaries, the viscosities and densities of the fluids, and the layer depths. A large number of experimental studies using a variety of materials have been performed. For the experiments using the most viscous fluids (e.g., silicone putties), the effective gravity was enhanced up to 800g by spinning in a centrifuge (e.g., Jackson and Talbot, 1989; Ramberg, 1967, 1968). Early experiments with putty and other non-Newtonian fluids have been extensively photographed and compared with geologic formations such as salt domes (Nettleton, 1934, 1943; Parker and McDowell, 1951; Ramberg, 1967, 1972). However, there was no comparison between these laboratory experiments and theory due to the unknown rheology of the laboratory materials. We restrict the discussion in the succeeding text to experiments with Newtonian fluids, easier to control.

Whitehead and Luther (1975) performed experiments under normal gravity and showed that the morphology of the rising pockets of light fluid depends on the viscosity ratio between the two fluids: more viscous instabilities develop in fingerlike 'diapirs,' while less viscous instabilities develop into mushroom-shaped 'cavity plume' (**Figure 20(a)** and **20(b)**). This latter case is of particular interest for the initiation of instabilities deep in the mantle since hot mantle material is probably also less viscous (e.g., Stacey and Loper, 1983). The characteristic spacing λ and growth rate σ_{RT} of RTI strongly depend on the viscosity of the materials (Bercovici and Kelly, 1997; Canright and Morris, 1993; Ribe, 1998; Selig, 1965; Whitehead and Luther, 1975), as well as on the geometry of their source (Kerr and Lister, 1988; Kerr et al., 2008; Lister and Kerr, 1989a,b; Namiki et al., 2013; Wilcock and Whitehead, 1991). For a thin denser lower layer of thickness h_d and viscosity ν_d underlying an infinite layer of viscosity ν_m, they scale as

$$\lambda = h_d F(\gamma) \qquad [5]$$

$$\sigma_{RT} = \frac{\Delta\rho g h_d}{\rho\nu} G(\gamma) \qquad [6]$$

where g is the gravity acceleration, $\gamma = \nu_d/\nu_m$, and $\nu = \max(\nu_d, \nu_m)$. Equation [6] shows that the instability growth rate is limited by the fluid with the larger viscosity. The functions F and G depend on the viscosity ratio and the geometry of the system. For a thin plane more viscous layer, linear stability analysis shows that $F(\gamma) \sim \gamma^{1/5}$ and $G(\gamma) \sim 1$, while for a thin plane less viscous layer, $F(\gamma) \sim \gamma^{-1/3}$ and $G(\gamma) \sim \gamma^{1/3}$ (Selig, 1965; Whitehead and Luther, 1975). Laboratory measurements agree with these scalings (**Figure 21**), which also confirms the ability of the linearized equations to predict the dominant wavelength of the developed instability (Whitehead and Luther, 1975). However, for instabilities developing from a cylinder in a denser fluid (**Figure 20(c)** and **20(d)**), the RTI spacing and growth rate still depend on the cylinder size but are independent of the viscosity ratio (**Figure 21**; Kerr and Lister, 1988; Lister and Kerr, 1989a,b). Lister et al. (2011) extended this study to the inclined buoyant viscous cylinder and showed that disturbances could propagate out of the system without significant growth for an angle >40°. Moreover, the presence of thermal diffusion can further delay the growth of the RTI (Kerr et al., 2008).

A related problem to the cylinder one is the rise and instability of a vertical sheet of lighter fluid. The upper front of the latter is seen to first destabilize in an array of diapirs. However, the rising diapirs are followed by a continuous, uniform sheet of buoyant

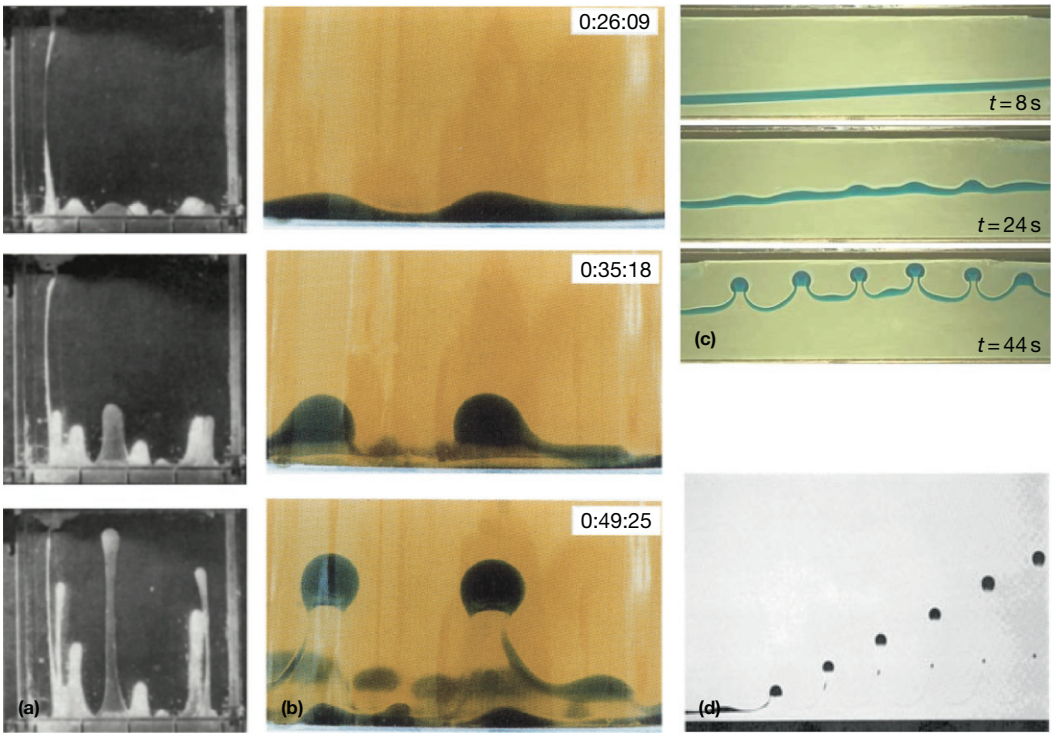


Figure 20 Development of Rayleigh–Taylor instabilities. (a) Fingers develop when the instabilities are more viscous than the ambient fluid (reproduced from Whitehead JA and Luther DS (1975) Dynamics of laboratory diapir and plume models. *Journal of Geophysical Research* 80: 705717). (b) Cavity plumes when the instabilities are less viscous (reproduced from Bercovici D and Kelly A (1997) Nonlinear initiation of diapirs and plume heads. *Physics of the Earth and Planetary Interiors* 101: 119–130).(c) Development of RTI out of a sloping cylinder (adapted from Kerr RC, Mériaux C, and Lister JR (2008) Effect of thermal diffusion on the stability of strongly tilted mantle plume tails. *Journal of Geophysical Research* 113: B09401. http://dx.doi.org/10.1029/2007JB005510). (d) Ascent of a cavity plume chain behind a steadily moving source (reproduced from Olson P and Singer H (1985) Creeping plumes. *Journal of Fluid Mechanics* 158: 511531).

fluid (Kerr and Lister, 1988). Close-up on the front shows that the buoyant fluid accumulates at the top of the sheet to form a buoyant cylinder. The gravitational instability then divides the cylinder into several plume heads (Namiki et al., 2013).

For a thin layer of light fluid embedded in denser fluid, linear stability again predicts a fastest growing wavelength much greater than the thickness of the low-viscosity layer, which agrees with the experimental results (**Figure 21**; Lister and Kerr, 1989a,b; Wilcock and Whitehead, 1991). Moreover, the displacement of the lower interface can reach a significant fraction of that of the upper interface. This leads to entrainment of the substratum into the light diapir (Ramberg, 1972; Wilcock and Whitehead, 1991). For very thin layers, a second instability is also observed at a scale much greater than the characteristic wavelength; it arises from perturbations predominantly involving thickening rather than translation of the buoyant layer (Wilcock and Whitehead, 1991).

Bercovici and Kelly (1997) described both theoretically and experimentally the evolution of the interface from the initial linear instability to the formation and liftoff of cavity plumes. This involves nonlinear feedback mechanisms between the growth of the instability and the draining of the low-viscosity channel, whereby the protoplume can stall for a long period of time before it separates and begins its ascent. The radii of the cavity plume a and of its trailing conduit R now scale as

$$a \sim h_d \gamma^{-2/9};\ R \sim h_d \qquad [7]$$

respectively, which shows that the radius of the head will be significantly larger than the conduit radius.

After RTI liftoff, the plumes may not always continue to rise along a vertical line as solitary bodies. Depending on their relative size, and the vertical and horizontal separation between them, they can become attracted to one another, clustering and even coalescing (Kelly and Bercovici, 1997). Combining theoretical, numerical, and laboratory results, Manga (1997) further suggested that attraction between plumes may be enhanced by plumes ability to deform (see Ribe, **Chapter 7.04**).

Once the instabilities have started, the light fluid will gather at the top of the box, where it will remain since the density configuration is now stable. An initially unstable compositional stratification will therefore go unstable only once. In that respect, the RTI appears as an essential process of segregation of two intermingled materials.

But the 'one-shot' character of RTI instability can be altered when the lighter fluid is continuously released from the bottom through a diffusive interface (Jellineck et al., 1999; Loper et al., 1988; Puthenveettil and Arakeri, 2005) or through melting (Jellineck et al., 1999). By introducing a silk membrane on the interface between a thin water layer and a heavy viscous

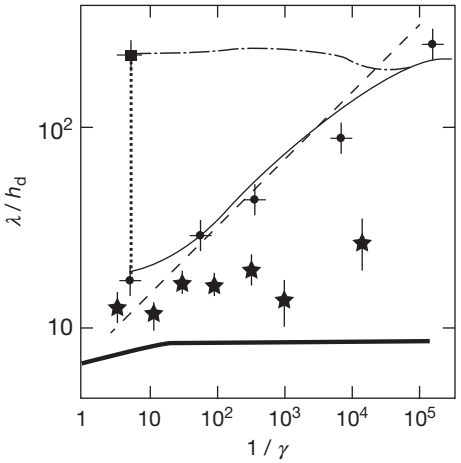


Figure 21 RTI wavelength normalized by the thin layer depth as the inverse of the viscosity ratio. The dashed line represents the infinite plane layer solution (eqn [6]). Dots are experimental data for a thin layer embedded in more viscous fluid; the dotted line connects observations from an experiment exhibiting a bimodal wavelength instability; characteristics of wavelength predicted by the numerical model are shown in dot-dashed and thin solid lines (reproduced from Wilcock WSD and Whitehead JA (1991) The Rayleigh-Taylor instability of an embedded layer of low-viscosity fluid. *Journal of Geophysical Research* 96: 12193–12200). Stars represent observations for a cylinder and the thick solid line shows the theoretical prediction (reproduced from Kerr RC and Lister JR (1988) Island arc and mid-ocean ridge volcanism, modelled by diapirism from linear source regions. *Earth and Planetary Science Letters*, 88: 143–152).

corn syrup layer, Loper and McCartney (1986) and Loper et al. (1988) observed intermittent RTI generation. When the lighter fluid is less viscous, the RTI morphology is 3-D as already described, whereby it is sheetlike when the lighter fluid is more viscous (Jellineck et al., 1999). Moreover, for high release flux (or high Reynolds number), the RTI can entrain on its way up a significant amount of the ambient denser fluid, and the final stratification is significantly altered by mixing compared to the initial stage (Jellineck et al., 1999).

More commonly, episodicity can be sustained indefinitely when the density field is diffusing. This is the case in thermal convection, to which we turn now.

7.03.5 Simple Rayleigh–Bénard Convection Studies

In buoyancy-driven flows, the exact governing equations are intractable. Some approximation is needed and the simplest one that admits buoyancy is the Oberbeck–Boussinesq approximation (see Ricard, Chapter 7.02). It assumes that

(a) in the equations for the rate of change of momentum and mass, density variations may be neglected except when they are coupled to the gravitational acceleration in the buoyancy force (in practice, it requires $\alpha\Delta T \ll 1$),
(b) all other fluid properties can be considered as constant over the experimental cell,
(c) viscous dissipation is negligible.

In this section, we shall focus on cases where the Oberbeck–Boussinesq approximation is valid. Since mantle material has a high *Pr*, emphasis will be on results using high-*Pr* fluids. The convective flows have usually been determined using planforms visualized by shadowgraph, heat flow measurements, or time series of temperature measurements within the experimental cell.

7.03.5.1 Convection at Relatively Low *Ra* ($Ra_c \leq Ra \leq 10^6$)

As predicted by linear stability theory (Chandrasekhar, 1961; Rayleigh, 1916), convection should set in above a critical value Ra_c, which depends on the boundary conditions (see Ribe, Chapter 7.04). Figure 22 shows a regime diagram of thermal convection as a function of Rayleigh and Prandtl numbers when both the top and bottom boundaries are rigid (zero velocity on the boundary) and isothermal. Rayleigh–Bénard convection exhibits a sequence of transitions toward chaos as *Ra* increases. These transitions can also be seen on the measurements of the heat flow Q_S extracted by the system (Figure 23), compared to the conductive heat flux $k\Delta T/H$. Their ratio defines the Nusselt number:

$$Nu = \frac{Q_s}{k.\Delta T/H} \qquad [8]$$

It is equal to unity for conduction and exceeds unity as soon as convection starts. It was by measuring Q_S as a function of ΔT that Schmidt and Milverton (1935) determined experimentally for the first time the critical Rayleigh number for the onset of convection. They confirmed the predicted theoretical value of 1708 with an accuracy better than 10%. Among others, Silveston (1958) extended the measurements to a wider range of fluids (Figure 23) and showed that, as predicted, the value of Ra_c does not depend on *Pr*. However, this is not the case for the subsequent transitions between patterns at higher *Ra* (Figure 22), for which there is a clear difference between low-*Pr* fluids (e.g., water, air, mercury, and compressed gases like helium) and fluids with $Pr \geq 100$.

For $Pr \geq 60$, laboratory experiments performed with silicone oils revealed the following sequence of patterns in the convective planform as *Ra* increases (Busse and Whitehead, 1971; Krishnamurti, 1970a,b; Richter and Parsons, 1975): 2-D rolls (Figure 24(a)) are stable only for Rayleigh numbers < 13 times critical. For *Ra* greater than this, a second set of rolls perpendicular to the original ones grows, and the new planform is rectangular. In this so-called bimodal regime (Figure 24(b)), the upwellings assume the geometry of two adjacent sides of a square, while the downwellings form the other two sides of the square. At $Ra \sim 100\ Ra_c$, a new planform develops through a collective instability: the corners of the bimodal rectangles join, and the resulting planform is time-dependent and resembles spokes radiating out of the central upwellings or downwellings (multimodal regime, Figure 24(c)). The two transitions in patterns are correlated with kinks in the *Nu*–*Ra* curves (e.g., Krishnamurti, 1970a,b; Malkus, 1954; Willis and Deardorff, 1967), and the *Ra*-value at each transition does not vary systematically with *Pr* (Busse and Whitehead, 1971; Krishnamurti, 1970a,b). However, the transition toward the time-dependent spoke pattern is not sharp, but there is a

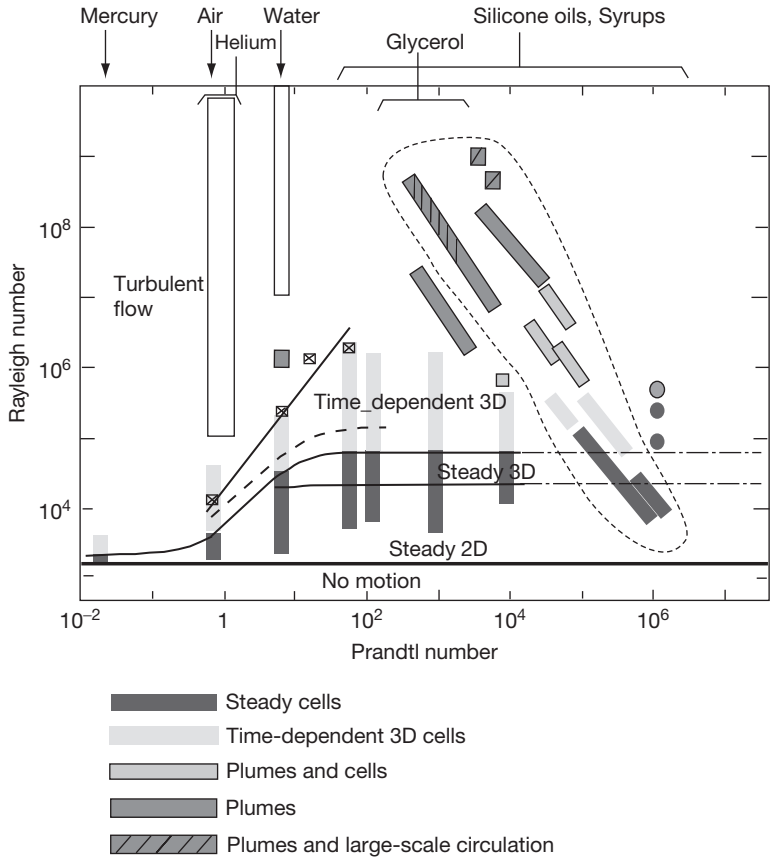


Figure 22 Regimes diagram of thermal convection. Data compiled from Krishnamurti (1970a,b, 1979), Whitehead and Parsons (1978), Nataf et al. (1984), Guillou and Jaupart (1995), Zhang et al. (1997), Manga and Weeraratne (1999), and Xi et al. (2004). The square bars show the range of experiments performed. The experiments performed with a viscosity contrast within the tank >10 are enclosed in the dotted line. The three round dots show experiments in centrifuge by Nataf et al. (1984).

domain in *Ra* where both time-dependent and stable cells can coexist. Hence, the pattern will depend strongly on the initial conditions. Under controlled initial conditions, Whitehead and Parsons (1978) observed stationary bimodal convection up to $Ra = 7.6 \times 10^5$, but an oscillating spoke pattern if starting from random conditions. At $Pr = 10^6$ (using a centrifuge), Nataf et al. (1984) observed steady motions at $Ra = 0.95 \times 10^5$, but time-dependent instabilities at $Ra = 6.3 \times 10^5$. When the Oberbeck–Boussinesq approximation applies, cold and hot currents have symmetrical characteristics, even for time-dependent spokes. And the horizontally averaged thermal structure is symmetrical relative to the temperature at the mean depth (**Figure 25**).

For intermediate and low *Pr*, the pattern for a Boussinesq fluid always consists of straight rolls above but close to onset. But the bifurcation toward stationary bimodal flow is replaced by an oscillatory secondary instability (e.g., Busse, 1978, 1989) and quite rapidly by chaos and then turbulence, as *Ra* increases further (**Figure 23**).

7.03.5.2 Patterns and Defects

The study of convection planforms and pattern selection is a very rich and fundamental field in itself and has motivated a large body of theoretical (see Ribe, **Chapter 7.04**) and experimental work. Reviews on the subject include papers by Busse (1978, 1989), Geitling (1998), Bodenschatz et al. (2000), and Ahlers (2005). Experiments are usually done in thin fluid layers with large aspect ratios (up to 100) and visualized by shadowgraph. We will mention here only a few striking results.

7.03.5.2.1 Generation of an initial prescribed pattern

To determine the domain of stability of the different patterns and follow the formation and evolution of the defects, the convective pattern must be initiated in a well-controlled way. The method developed by Chen and Whitehead (1968) consists in placing a grid made up of alternating blocked and clear areas over the top transparent channel. A 300–500 W incandescent lamp then is shone down through the pattern, so that the fluid lying below is slightly radiatively heated in the desired pattern. After a certain time (~1 h), the temperature of the top bath is decreased and the temperature of the bottom bath increased at equal rates (typically $2^\circ\,\mathrm{min}^{-1}$) until the desired Rayleigh number is reached. It is important to change the baths at the same rates in order to exclude asymmetrical or hexagonal modes of flow. Convection then starts with the location of the rising limbs of the cells being controlled by the places that were heated.

7.03.5.2.2 Patterns for Pr ≥ 100

Chen and Whitehead (1968) and Busse and Whitehead (1971) studied experimentally the stability of rolls for a silicone oil with

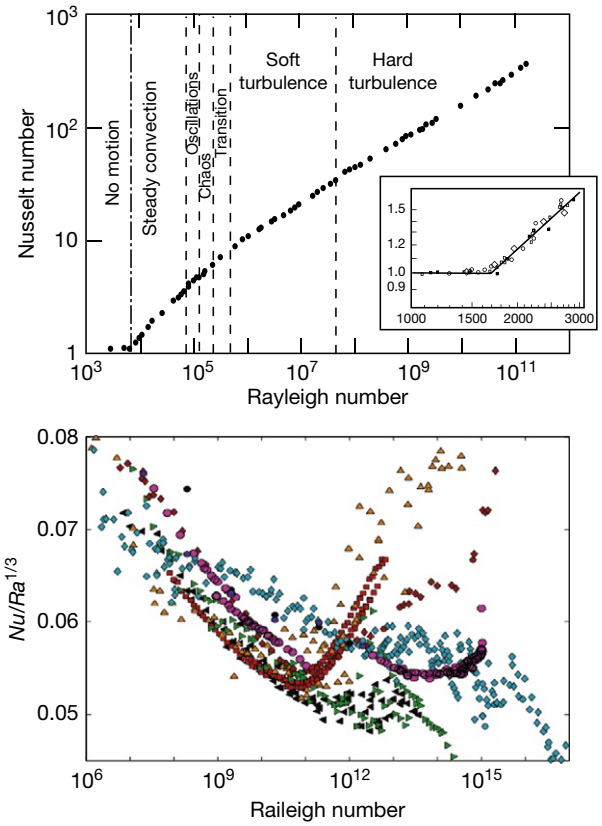


Figure 23 (top) Nusselt number as a function of Rayleigh number measured in He. The different transitions observed are indicated (adapted from Heslot F, Castaing B, and Libchaber A (1987) Transitions to turbulence in helium gas. *Physical Review A* 36: 5870–5873). The insert shows the data from Silveston (1958) around the convection onset. (bottom) Heat transport as a function of the Rayleigh number in a compensated power law plot for experiments and numerical simulations. For aspect ratio $\gamma = 1/2$ in helium, green triangles pointing to the right are for experiments run in Chicago, cyan diamonds stand for Oregon, and orange triangles pointing upward for Grenoble. For $\gamma = 1$, red squares stand for Grenoble experiments (in He), black triangles pointing to the left are for Brno (in He), brown diamonds for Trieste (in He), and magenta circles for Gottingen (in SF_6). Blue circles are for numerical simulations at $Pr = 0.7$ and $\gamma = 1/2$; black circles for $Pr = 2$ and $\gamma = 0.23$. The data are all corrected for wall finite conductivity. Adapted from Chilla F and Schumacher J (2012) New perspectives in turbulent Rayleigh–Bénard convection. *The European Physical Journal E* 35: 58–83.

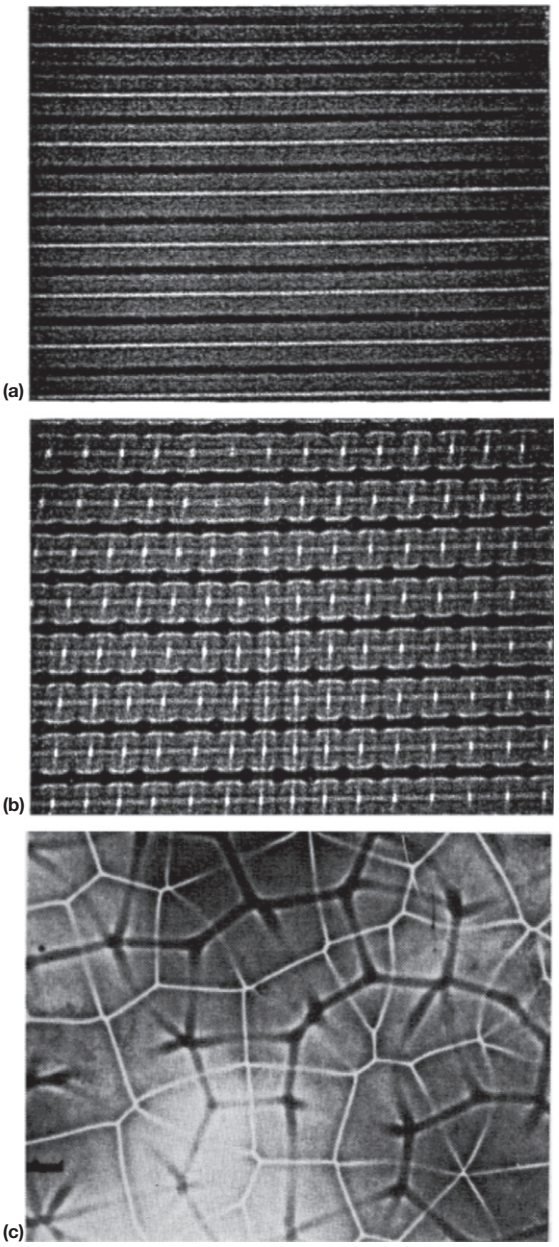


Figure 24 Shadowgraphs of the different stable cellular patterns in isoviscous convection at high *Pr*: (a) rolls, $Ra = 20 \times 10^3$; (b) bimodal, $Ra = 64 \times 10^3$ (reproduced from Busse FH and Whitehead JA (1971) Instabilities of convection rolls in high Prandtl number fluid. *Journal of Fluid Mechanics* 47: 305–320); (c) spoke, $Ra = 1.4 \times 10^5$ (reproduced from Richter FM and Parsons B (1975) On the interaction of two scales of convection in the mantle. *Journal of Geophysical Research* 80: 2529–2541).

$Pr = 100$. Rolls of known aspect ratio (roll width over tank thickness) were initiated and allowed to evolve. The results were compared with the theory of Busse (1967a,b), which predicts a balloon-shaped region of stable rolls (Figure 26). The zigzag instability (Figure 27(a)) occurs when the aspect ratio is >1 and tends to reduce it to 1. The cross roll instability (Figure 27(b)) occurs when the aspect ratio is much <1 or when the Rayleigh number exceeds about 2×10^4: the initial rolls are replaced by two sets of rolls with different wavelength at right angles to each other. The pattern then evolves toward one set of rolls with increased aspect ratio (in the first case) or to the bimodal mode (in the second case). When two sets of rolls of different wavelengths are combined, local rearrangement occurs through the pinching instability (Figure 27(c)). The agreement between the experimental results and the theory is quite remarkable (Figure 26; see also Ribe, Chapter 7.04).

7.03.5.2.3 Patterns for Pr < 100

Patterns and defects have been extensively studied more recently, using primarily compressed gases as the working fluid (small cell sizes can be used and *Pr* can be tuned between 0.17 and 30), sensitive shadowgraph visualization coupled to digital image

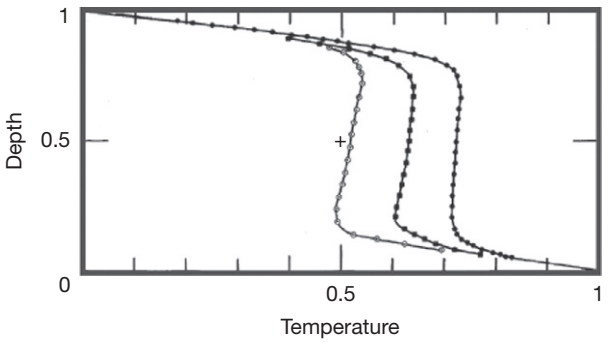


Figure 25 Nondimensional temperature as a function of depth. Circles, Silicone oil, $\gamma = 3$, $Ra = 2.16 \times 10^5$; squares, L-100, $\gamma = 22$, $Ra = 1.36 \times 10^5$; filled discs, L-100, $\gamma = 750$, $Ra = 4.95 \times 10^5$. In all but the last case, data near the boundaries have been omitted. Reproduced from Richter FM, Nataf H-C, and Daly SF (1983) Heat transfer and horizontally averaged temperature of convection with large viscosity variations. *Journal of Fluid Mechanics* 129: 173–192.

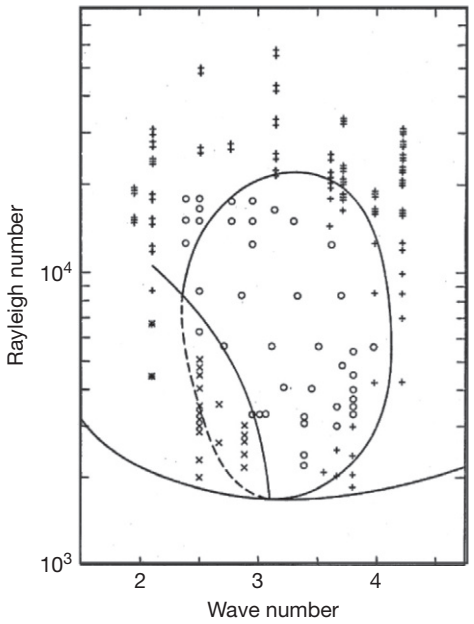


Figure 26 Domain of stability of rolls as a function of the wave number and the Rayleigh number ('Busse's balloon'). Solid lines are the theoretical predictions (Busse, 1978). Experimental results by Busse and Whitehead (1971): o, stable rolls; x, zigzag instability; +, cross roll instability leading to rolls; ++ cross roll instability leading to bimodal convection; +++ cross roll instability inducing transient rolls with subsequent local processes. Reproduced from Busse FH and Whitehead JA (1971) Instabilities of convection rolls in high Prandtl number fluid. *Journal of Fluid Mechanics* 47: 305–320.

analysis, and quantitative heat flow measurements (for a review, see Bodenschatz et al., 2000). **Figure 27** shows examples of the patterns observed. Above but close to onset, the pattern for a Boussinesq fluid always consists of straight rolls, possibly with some defects induced by the sidewalls. At $Pr \sim 1$, the skewed-varicose instability arises to transform rolls into rolls of larger wavelength (**Figure 27(d)**). When non-Boussinesq conditions prevail, hexagons develop instead of rolls (**Figure 27(e)**). Further above onset, spiral-defect chaos (Morris et al., 1993) occurs in systems with Pr of order 1 or less (**Figure 27(f)**). This state is a bulk property and is not induced by sidewalls. On the other hand, spirals occur in circular cells (**Figure 27**).

7.03.5.3 Thermal Convection at High Rayleigh Numbers ($10^5 \leq Ra$)

At high Ra, the convective pattern becomes disorganized, and flow is generated locally by TBL instabilities (Elder, 1968), which develop into plumes (**Figures 9** and **10(c)**). Their thermal anomalies are easily recorded by thermocouples located within the fluid (**Figure 28**; e.g., Castaing et al., 1989; Davaille and Jaupart, 1993; Deardorff et al., 1969; Lithgow-Bertelloni et al., 2001; Sparrow et al., 1970; Townsend, 1957; Weeraratne and Manga, 1998).

7.03.5.3.1 TBL instabilities

Plume generation in fully developed convection is a cyclic process in which the TBLs grow by thermal diffusion, become unstable when they reach a critical thickness δ_c such that $Ra(\Delta T,\ \delta_c) = Ra_c$, and then empty themselves rapidly into plumes, at which point the cycle begins again (**Figure 29**). The characteristic timescale τ_c for this process is the time required for the growing TBL to become unstable and is (Howard, 1964)

$$\tau_c = \frac{H^2}{\pi\kappa}\left(\frac{Ra_c}{Ra(\Delta T, H)}\right)^{2/3} \qquad [9]$$

Because $Ra \sim H^3$ (eqn [1]), τ_c is independent of the layer depth H: the two TBLs do not interact anymore. This phenomenological model (Howard, 1964) and the scaling [9] are confirmed by laboratory experiments (**Figures 29** and **30**; e.g., Androvandi et al., 2011; Davaille, 1999a; Davaille and Jaupart, 1993; Davaille and Vatteville, 2005; Manga et al., 2001; Sparrow et al., 1970). Experiments further show that the typical spacing between plumes is three to six times δ_c (e.g., Androvandi et al., 2011; Asaeda and Watanabe, 1989; Davaille et al., 2002; Jellinek and Manga, 2004; Sparrow et al., 1970; Tamai and Asaeda, 1984). Moreover, the plumes seem connected by a network of ridges near the base of the layer (Androvandi et al., 2011; Asaeda and Watanabe, 1989; Puthenveettil and Arakeri, 2005; Tamai and Asaeda, 1984; Zhou et al., 2007). This morphology was confirmed by detailed numerical studies (e.g., Christensen and Harder, 1991; Houseman, 1990; Parmentier and Sotin, 2000; Tackley, 1998; Trompert and Hansen, 1998a). Plumes have a head-and-stem structure, and both head-and-stem diameters decrease with increasing Ra (Lithgow-Bertelloni et al., 2001). Plumes can eventually detach from the TBL once they have drained it (**Figure 29**), as was observed for RTI (e.g., Bercovici and Kelly, 1997; **Figure 20**). They are transient features with a lifetime comparable to τ_c (Davaille and Vatteville, 2005). Since the two TBLs do not interact, the global heat flow should not depend on H anymore, implying

$$Nu \sim Ra^{1/3} \qquad [10]$$

This scaling was verified by Globe and Dropkin (1959) and Guillou and Jaupart (1995) using silicone oils for Ra up to $Ra \sim 2 \times 10^7$. Goldstein et al. (1990), who employed an electrochemical analogue to convection to reach $Ra \sim 10^{12}$ with

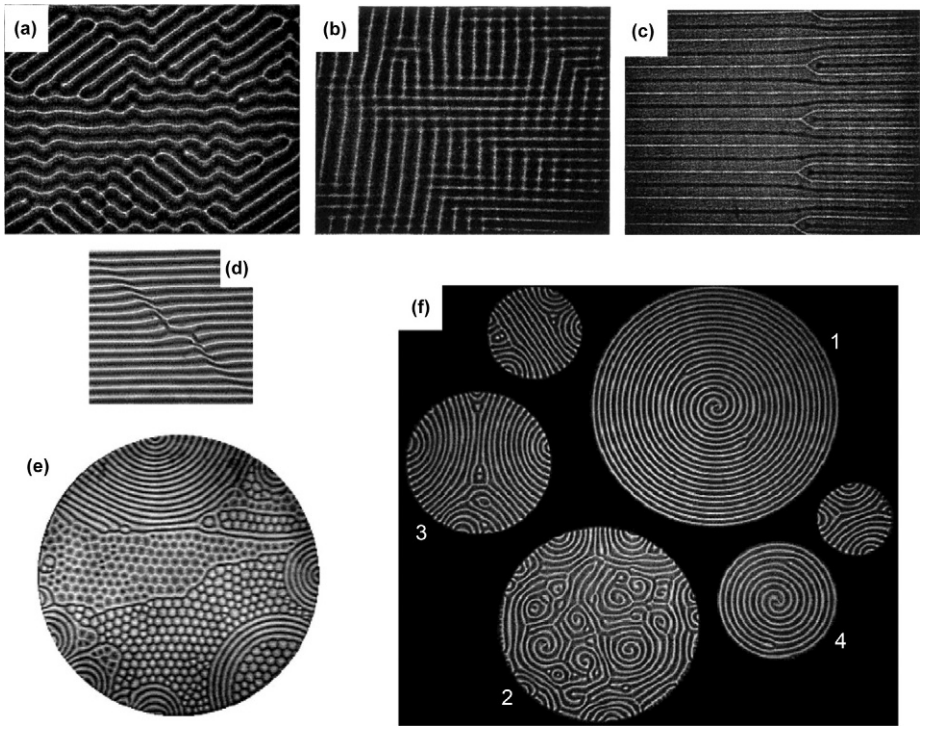


Figure 27 Patterns and defects. (a) Zigzag, (b) cross roll, (c) pinching instabilities (reproduced from Busse FH and Whitehead JA (1971) Instabilities of convection rolls in high Prandtl number fluid. *Journal of Fluid Mechanics* 47: 305–320); (d) skewed-varicose instability (reproduced from Plapp BB (1997) Spiral-pattern formation in Rayleigh–Bénard convection. PhD Thesis. Cornell University, Ithaca, NY); (e) coexisting up- and downflow hexagons together with rolls (reproduced from Assenheimer M and Steinberg V (1993) Rayleigh–Bénard convection near the gas–liquid critical point. *Physical Review Letters* 70: 3888–3891); (f) two-armed spiral patterns in cells 1 and 4. Spiral-defect chaos in cell 2 (reproduced from Plapp BB, Egolf DA , Bodenschatz E, and Pesch W (1998) Dynamics and selection of giant spirals in Rayleigh–Bnrad convection. *Physical Review Letters* 81: 5334–5337).

$Pr = 2750$, recovered also eqn [10]. This regime of plumes has sometimes been named soft turbulence.

7.03.5.3.2 Soft and hard turbulence at low Pr

Kraichman (1962) had predicted the occurrence of an 'ultimate regime' based on boundary layer analysis, with an increase of the heat transfer in $Ra^{1/2}$ at very high Ra. Steady-state experiments by Heslot et al. (1987) and Castaing et al. (1989) in helium ($0.6 \leq Pr \leq 1.7$) indicated that the spectrum of temperature fluctuations changes above $Ra = 4 \times 10^7$, defining a new dynamic regime called 'hard turbulence.' This transition is also seen on the Nu–Ra relationship (**Figure 23**, top), which follows the 1/3 power law eqn [10] for $Ra \leq 4 \times 10^7$ (the 'soft turbulence' regime) but changes to a 2/7 power law for $Ra \geq 4 \times 10^7$. On the other hand, Katsaros et al. (1977) recovered the 1/3 power law eqn [10] for Ra up to 10^9 during the transient cooling from above of a water tank. Castaing et al. (1989) suggested that, in steady state, the dynamics of boundary layer instabilities were affected by the establishment of a large-scale circulation over the whole tank. Since the experiments were done in a container with aspect ratio 1, the large-scale circulation could have been caused by the wall effects. However, the establishment of a large-scale circulation had already been observed by Krishnamurti and Howard (1981) for larger aspect ratio and is well known in atmospheric convection. Much theoretical and experimental work has been done since to confirm or disprove, extend, and explain the new regime (for recent reviews, see Ahlers et al., 2009; Busse, 2012; Chilla and Schumacher, 2012; Lohse and Xia, 2010; Roche et al., 2010). Following Kraichman (1962), Grossmann and Lohse (2000, 2011) proposed a systematic theory to predict all the different asymptotic regimes as a function of Ra and Pr. In the search for the ultimate turbulent regime, $Ra \sim 10^{17}$ has now been obtained using cryogenic He (Niemela et al., 2000). For aspect ratios smaller than 1, several regimes seem to coexist with very different heat transfer (**Figure 23**, bottom). This could be due to the geometry of the large-scale circulation, which can switch from one cell to several superimposed cells in narrow tanks (e.g., Chilla et al., 2004), or to the nature (finite thermal conductivity or rugosity) of the tank horizontal and vertical boundaries (e.g., Roche et al., 2010). More work needs now to be done for aspect ratios larger than 1.

Although this Ra range is not relevant for solid-state mantle convection, it is interesting for convection in magma oceans during the early stages of the Earth. In the latter case, rotation will also play an important role (e.g., King et al., 2013).

7.03.5.3.3 Large-scale (cellular) circulation and plumes at high Pr

To obtain high Rayleigh number in high-Pr fluids requires large temperature differences (typically $\geq 20\,^{\circ}C$) and large depths (typically $\geq$20–30 cm). The aspect ratios of these

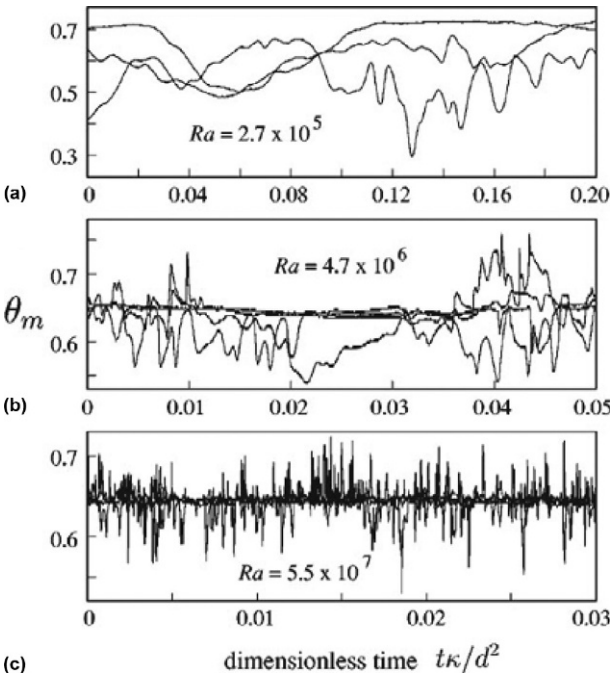


Figure 28 Temperature fluctuations $\theta_m = (T - T_{top})/\Delta T$ at three thermocouples located at midheight in a tank of convecting corn syrup. Time is normalized by the diffusive timescale h^2/κ. (a) Unsteady convection dominated by large-scale flow ($Ra = 2.5 \times 10^5$). (b) Increased plume activity along with large-scale flow ($Ra = 4.7 \times 10^6$). (c) Plume-dominated convection with short-period fluctuations ($Ra = 5.5 \times 10^7$). Reproduced from Weeraratne D and Manga M (1998) Transitions in the style of mantle convection at high Rayleigh numbers. *Earth and Planetary Science Letters* 160: 563568.

experiments are therefore small (between 1 and 3 at most). Even for fluids with very weak temperature dependence of viscosity (e.g., silicone oils), the viscosity can no longer be treated as constant (**Figures 2** and **22**). Its value is therefore taken at the mean of the boundary temperatures $(T_{bot} + T_{top})/2$.

Weeraratne and Manga (1998), Manga and Weeraratne (1999), and Manga et al. (2001) determined the different styles of time-dependent convection from temperatures recorded on the tank midplane (**Figure 28**). For steady flow ($Ra \leq 10^5$), these remain constant. When the flow first becomes unsteady, temperature fluctuations with large amplitude and long period appear (**Figure 28(a)**). As Ra increases, the amplitude of long-period fluctuations decreases and small short-period fluctuations appear (e.g., $Ra \sim 5 \times 10^6$; **Figure 28(b)**). At still higher Ra, the large-amplitude fluctuations disappear completely and only the short-period fluctuations remain (**Figure 28(c)**). As mushroom-shaped plumes were observed in this regime, the small-scale fluctuations were interpreted as their signature and the long-period signal as the signature of large-scale (cellular) flow. From $Ra \sim 10^7$ to 2×10^8, no large-scale circulation was observed. Such circulation was also absent from transient experiments where the fluid was continuously heated from below (Lithgow-Bertelloni et al., 2001) or cooled from above (Davaille and Jaupart, 1993), and the other boundaries were insulated. Using glycerol (Zhang et al., 1997, 1998) and dipropylene glycol (Xi et al., 2004) in aspect ratio 1 cells, a different picture was found for $Ra \sim 10^8$–10^9: a large-scale circulation develops that entrains the plumes as soon as they have formed. Measuring the velocity field by PIV, Xi et al. (2004) demonstrated that the flow becomes organized because of the plumes, which interact through their velocity boundary layers typically as soon as they have risen a distance comparable to their spacing: plume clustering then occurs, as in Rayleigh–Taylor instabilities (Bercovici and Kelly, 1997) and in experiments on the interaction between two thermal plumes (Moses et al., 1993). In this regime, where plumes and large-scale circulation coexist, the exponent of the Nu–Ra power law becomes again closer to 2/7 (Xi et al., 2004). It would now be interesting to run the same experiments with larger aspect ratio to determine if the large-scale circulation always spans the whole tank or if it has its own wavelength.

7.03.5.4 Convective Regime in an Isoviscous Mantle

From **Figure 22** and eqn [1], one can determine the convective regime of a mantle layer as a function of its thickness and average viscosity (**Figure 31**). Plumes will develop in the upper mantle (of thickness 660 km) only if its viscosity is lower than 10^{21} Pa s. For an average mantle viscosity of 10^{22} Pa s, plumes will be generated only if the mantle layer thickness exceeds 2000 km. Convective motions originating at the core mantle boundary (CMB) and developing over the whole mantle thickness should therefore take the form of plumes (Loper and Stacey, 1983; Parmentier et al., 1975). However, those plumes are transient phenomena. According to eqn [9], $\tau_c \sim 10$, 40, and 200 My for dynamic viscosities $\eta_m = 10^{19}$ Pa s, 5×10^{20} Pa s, and 10^{22} Pa s, respectively. Such recurrence times are probably also upper bounds on plume lifetime. The corresponding values of the critical TBL thickness $\delta_c \sim H{:}(Ra_c/Ra)^{1/3}$ for the same viscosities are 31, 62, and 140 km, respectively, which implies a plume spacing 100–840 km in the mantle. Given the 2900 km thickness of the mantle, plume clustering could occur (e.g., Schubert et al., 2004). Moreover, from **Figure 29**, we expect that a plume will eventually detach from the hot TBL from which it originated. This implies that the absence of a plume conduit in a tomographic image need not indicate the absence of a plume.

However, if isoviscous convection predicts plumes, it does not predict plates, which are the main signature of convection on the Earth's surface (e.g., Ricard, **Chapter 7.02**). Part of the difficulty in determining the convective pattern in the mantle resides in the complexity of mantle rheology, which depends on temperature (strongly), pressure, partial melting (strongly), water content (strongly), chemistry, and strain rate (see Kohlstedt, Volume 2). The rheology of the lithosphere is especially difficult to quantify since it comprises a brittle skin and a viscous lower part. The convective pattern driven by the cold upper mantle boundary appears to be cellular, although time-dependent. Inverting **Figure 31** for this regime, it would therefore indicate an effective mantle viscosity $> 10^{23}$ Pa s, that is, much lower than the viscosity of the cold lithosphere, but at least an order of magnitude higher than the measured bulk mantle viscosity. This asymmetry in mantle convection could be caused by its mixed heating mode (internal and bottom heating) and by its rheology, in particular the temperature dependence of viscosity. We now turn to the experiments that have been devoted to these two effects.

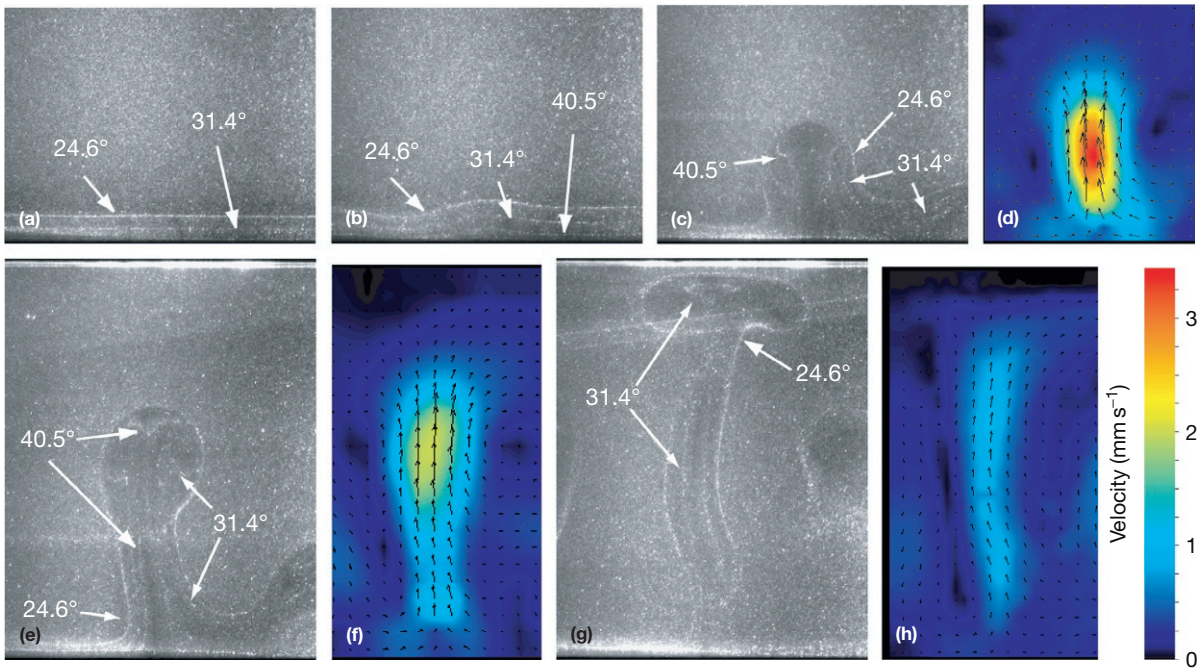


Figure 29 TBL instabilities in a layer of sugar syrup, initially at 21 °C and suddenly heated from below at 53 °C ($Ra = 1.7 \times 10^6$). (a) Negative of the picture taken at $t = 300$ s. The isotherms appear as white lines. The TBL is growing by conduction from the lower boundary. (b) $t = 400$ s. The TBL becomes unstable. (c) $t = 460$ s. A thermal plume, outlined by the 24.6 °C isotherm, rises from the TBL, and the TBL begins to shrink. (d) Corresponding velocity field deduced from PIV. The color background represents the velocity magnitude. (e) $t = 500$ s. The plume is well developed and the TBL is emptying itself in it. (f) Corresponding velocity field. (g) $t = 600$ s. The plume head has reached the upper boundary and begins to spread under it. The TBL has nearly disappeared and the conduit is disconnected from its source. (h) Corresponding velocity field. Reproduced from Davaille A and Vatteville J (2005) On the transient nature of mantle plumes. *Geophysical Research Letters* 32. http://dx.doi.org/10.1029/2005GL023029.

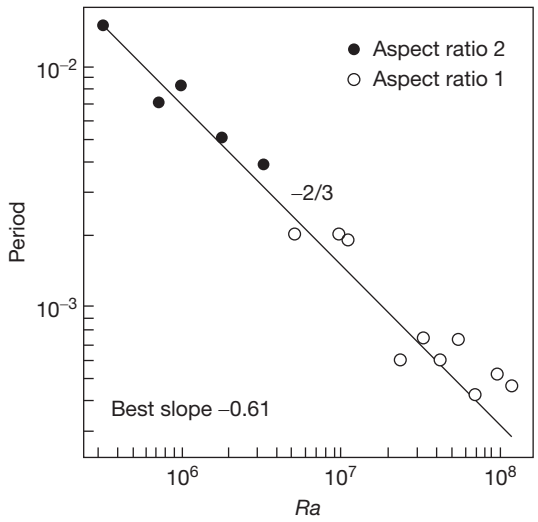


Figure 30 Periodicity of the TBL instabilities as a function of Rayleigh number. The solid line represents Howard's scaling in $Ra^{-2/3}$. From Manga M and Weeraratne D (1999). Experimental study of non-Boussinesq Rayleigh–Bénard convection at high Rayleigh and Prandtl numbers. *Physics of Fluids* 10: 2969–2976.

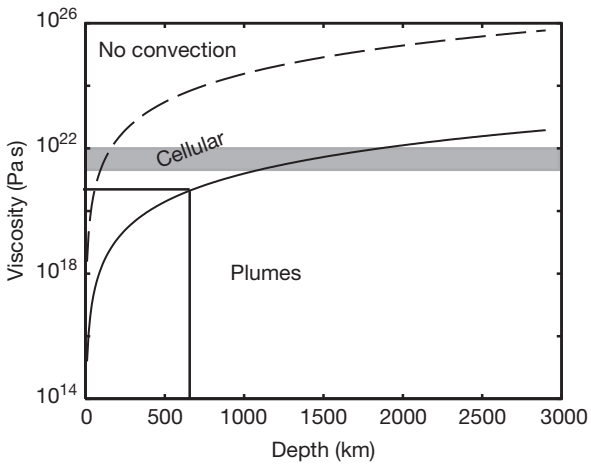


Figure 31 Convective regime developing in a mantle layer as a function of its thickness and viscosity. The pattern is cellular for $Ra \geq Ra_c = 650$, and in plumes for $Ra \geq 106$. In gray is the average viscosity inferred for the whole mantle (Kolhstedt, volume 2). The thin gray line delimits the upper mantle. The calculation has been done from eqn [1] taking $\kappa = 10^{-6}$ m^2 s^{-1}, $\alpha = 2 \times 10^{-5}$°K^{-1}, and $\Delta T = 3000$°K.

7.03.6 Temperature-Dependent Viscosity

A strong temperature dependence of the viscosity breaks the symmetry in the convection cell: cold instabilities are now more viscous than the fluid in the bulk of the tank, while the hot instabilities are less viscous. According to what has been observed in RTI (Section 7.03.4, Figure 20), their morphology should change. Figure 32 shows that this is indeed the case: hot TBL instabilities are mushroom-shaped, while cold instabilities become diapirs, which can fold on the bottom plate before spreading (Figure 32(b)). Material in the cold TBL also flows less easily and rapidly than in the rest of the fluid, which will diminish the efficiency of convection and the heat transport. In addition to *Pr* and *Ra*, the system is now also characterized by the viscosity ratio $\gamma = \eta(T_{top})/\eta(T_{bot})$, and we need to choose the reference viscosity in *Ra*, for example, at the mean of the boundary temperatures ($Ra_{1/2}$) or at the hot bottom temperature (Ra_{bot}).

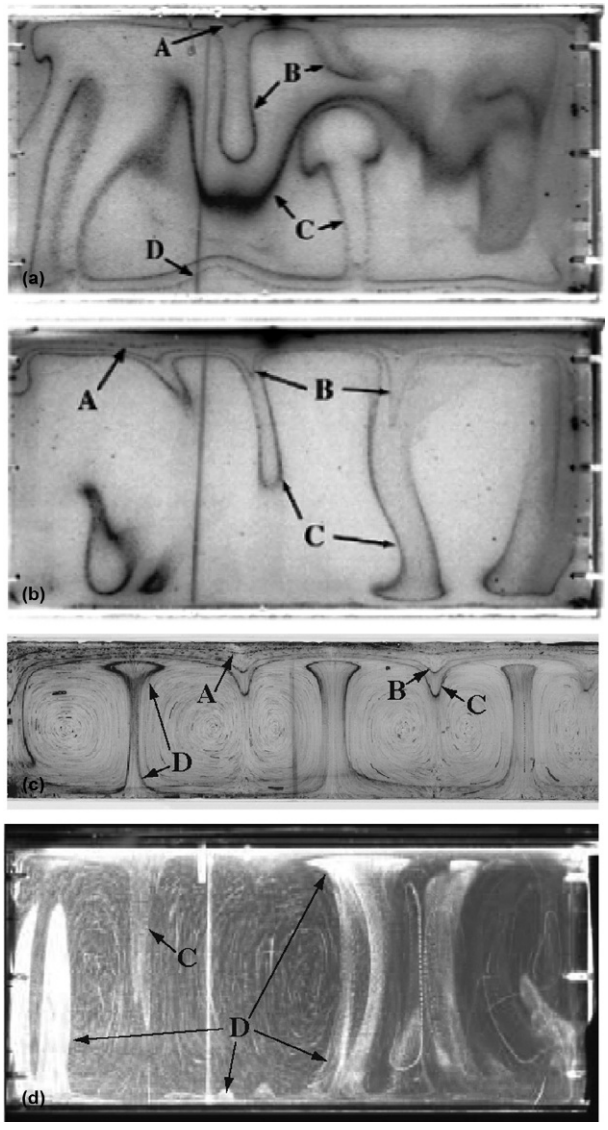


Figure 32 Thermal convection in a strongly temperature-dependent viscosity fluid (glucose syrup) cooled from above and heated from below. The black (a, b, c) or white (d) solid lines are the isotherms (A:10.5 °C; B:31.4 °C; C:24.6 °C; D:40.5 °C) (a) Asymmetry of the hot and cold TBL instabilities ($Ra_{bot} = 4{:}7 \times 10^6$; $\gamma = 116$). (b) Cold TBL for sugar syrup cooled from above only ($Ra_{bot} = 1{:}2 \times 10^7$; $\gamma = 1580$). (c) Time-lapse photography (over 30 s.) showing the particle trajectories as well as the isotherms at low *Ra* ($Ra_{bot} = 9{:}7 \times 10^4$; $\gamma = 888$); (d) Time-lapse photography (over 2000 s) at high Ra and intermediate γ showing the two scales of convection: several hot plumes are encaged in a 'box' delimited by stable cold downwellings ($Ra_{bot} = 5{:}3 \times 10^7$; $\gamma = 142$).

7.03.6.1 Rigid Boundaries

7.03.6.1.1 Onset of convection

Stengel et al. (1982) studied theoretically and experimentally the onset of convection for $\gamma \leq 3400$. Ra_c remains constant at low γ, increases at moderate γ to a maximum for $\gamma \sim 3000$, and then decreases. According to Stengel et al. (1982), this occurs because convection begins first in a (hotter) less viscous sublayer where the local Rayleigh number is maximum. Richter et al. (1983) increased the range of viscosity ratio to 10^5 and found similar results. Moreover, finite-amplitude disturbances are seen to grow at Rayleigh numbers below critical, as predicted by Busse (1967a,b) for fluids with weakly temperature-dependent viscosity.

7.03.6.1.2 Patterns and regimes

Figures 33 and 34 show the regime diagrams established from experimental, numerical, and theoretical analyses. Richter (1978) explored the stability of rolls, using glycerine and L100 (polybutene), for $Ra_{1/2} \leq 25 \times 10^3$ and $\gamma \leq 20$. For large γ, hexagons were found to be the stable pattern. White (1988) used golden syrup to study the stability of patterns for $\gamma \leq 1000$ and $Ra_{1/2} \leq 63 \times 10^3$ (Figure 34). Besides rolls and hexagons, a new planform of squares (Figure 34(b)) was found to be stable at large viscosity variations, and pattern instabilities developed in a mosaic form (Figure 34(e) and 34(f)). At fixed wave number, rolls are always unstable as soon as $\gamma > 40$. Moreover, squares break down to a spoke pattern around $Ra_{1/2} = 20\,000$, that is, much sooner than in the constant-viscosity case (transition at $Ra \sim 10^5$). The wavelength of squares is also observed to decrease with increasing γ. Temperature profiles through the layer (e.g., Figure 25) revealed that this shift is associated with the development of a thick, cold boundary layer and an increase of the interior temperature (Richter et al., 1983; White, 1988). The temperature profiles allowed the calculation of the convective heat flow and viscosity profile, which showed that most of the temperature variations across the tank occur in the cold TBL and that the top part of the latter becomes stagnant (i.e., zero convective heat flow; Figure 36). Hence, convection develops in a sublayer where the viscosity ratio reaches at most 100 (Richter et al., 1983). This stagnant lid regime has been observed for *Ra* up to 10^8 (Davaille and Jaupart, 1993).

For $Ra_{bot} \geq 10^5$, the convective pattern is determined by the local characteristics of the two TBLs (Androvandi et al., 2011; Figures 32(c), 32(d), and 33). So, at high Rayleigh numbers and intermediate viscosity ratios, three different scales of convection develop. The largest convective scale is cellular, with cold downwelling sheets of viscous fluid encasing hotter,

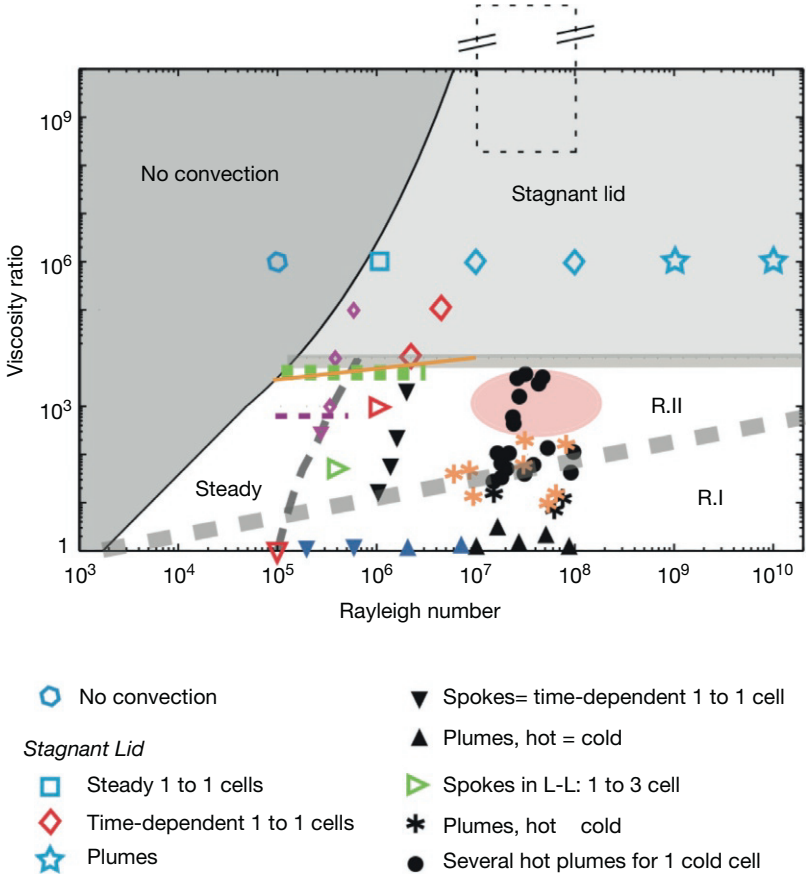


Figure 33 Convective pattern as a function of the Rayleigh number at the bottom boundary and the viscosity ratio for 3-D experiments in a Cartesian box. Only experiments allowing to determine the convective pattern have been considered. Open symbols designate experiments run with free surfaces (mostly numerical simulations) and solid symbols designate experiments run with solid boundaries. Black: Androvandi et al. (2011); light blue: Trompert and Hansen (1998a); red: Tackley et al. (1994), Tackley (1996), and Ratcliff et al. (1997); green: Weinstein and Christensen (1991); dark blue: Guillou and Jaupart (1995); orange: Schaeffer and Manga (2001). The dashed gray border between the steady and time-dependent whole-layer regimes is after White (1988). The thick dashed gray line shows the theoretical limit of Solomatov (1995) between the low-viscosity ratio regime ('R.I') and the 'transitional' or 'sluggish lid' regime ('R.II'). The thick solid gray line between the whole-layer ('R.II') and the stagnant lid regimes is after Solomatov (1995); the orange line is from the steady-state calculations of Ogawa et al. (1991); the dashed thick purple line from Nataf and Richter (1982) and Richter et al. (1983); and the green thick line after Giannandrea and Christensen (1993). The limit between no convection and convection is drawn after Stengel et al. (1982). The pink area designates the parameter range where the two scales of convection (cold cells encasing several hot plumes) are reminiscent of the present-day Earth's mantle convective pattern, while the thin black dashed line shows the actual parameter range for the Earth's mantle.

less viscous, parts of the system. Within each of those cells develop several (typically 3–7) hot 3-D upwelling plumes (**Figure 32(d)**). Upon impinging under the cold TBL, each plume in turn generates locally a small ring of cold material that does not reach the bottom of the tank (Androvandi et al., 2011). The wavelength of plumes and cells depends on the local properties (and Ra) of each TBL.

For $10^4 \leq Ra \leq 6 \times 10^5$ and $\gamma < 10^5$, the heat transfer differs little from that of a uniform-viscosity fluid when the Rayleigh number is defined with the viscosity at the mean of the boundary temperatures (Booker, 1976; Booker and Stengel, 1978; Richter et al., 1983). The relationship between Nusselt number and supercriticality ($Ra/Ra_c(\gamma)$) is even more remarkable since it is independent of γ and indistinguishable from the results of Rossby (1969) for constant viscosity:

$$Nu = 1.46\left(\frac{Ra}{Ra_c(\gamma)}\right)^{0.281} \qquad [11]$$

Subsequent high-quality heat flow measurements by Giannandrea and Christensen (1993) agree with this law. Weeraratne and Manga (1998), Manga and Weeraratne (1999), and Manga et al. (2001) extended the range of Ra up to 2×10^8 for $\gamma \leq 1800$. We already described in the previous section the sequence of regimes that they observed for $\gamma \leq 170$, that is, in absence of a conductive lid (**Figure 22**). Their heat flux data, measured using thermocouples, agree with eqn [11] for Ra up to 2×10^6 (**Figure 35**). For higher Ra and higher γ, the heat flux is less than that predicted by eqn [11] and by numerical calculations at infinite Pr. This trend was attributed to inertial effects (Manga and Weeraratne, 1999). However, subsequent heat flux measurements using TLC isotherms (Androvandi, 2009) for Ra_{bot} up to 10^8 and γ up to 4×10^3 recovered a law similar to eqn [11]. In the latter, the exponent $n = 0.281 < 1/3$ is coherent with the observations that the hot and cold TBLs interact.

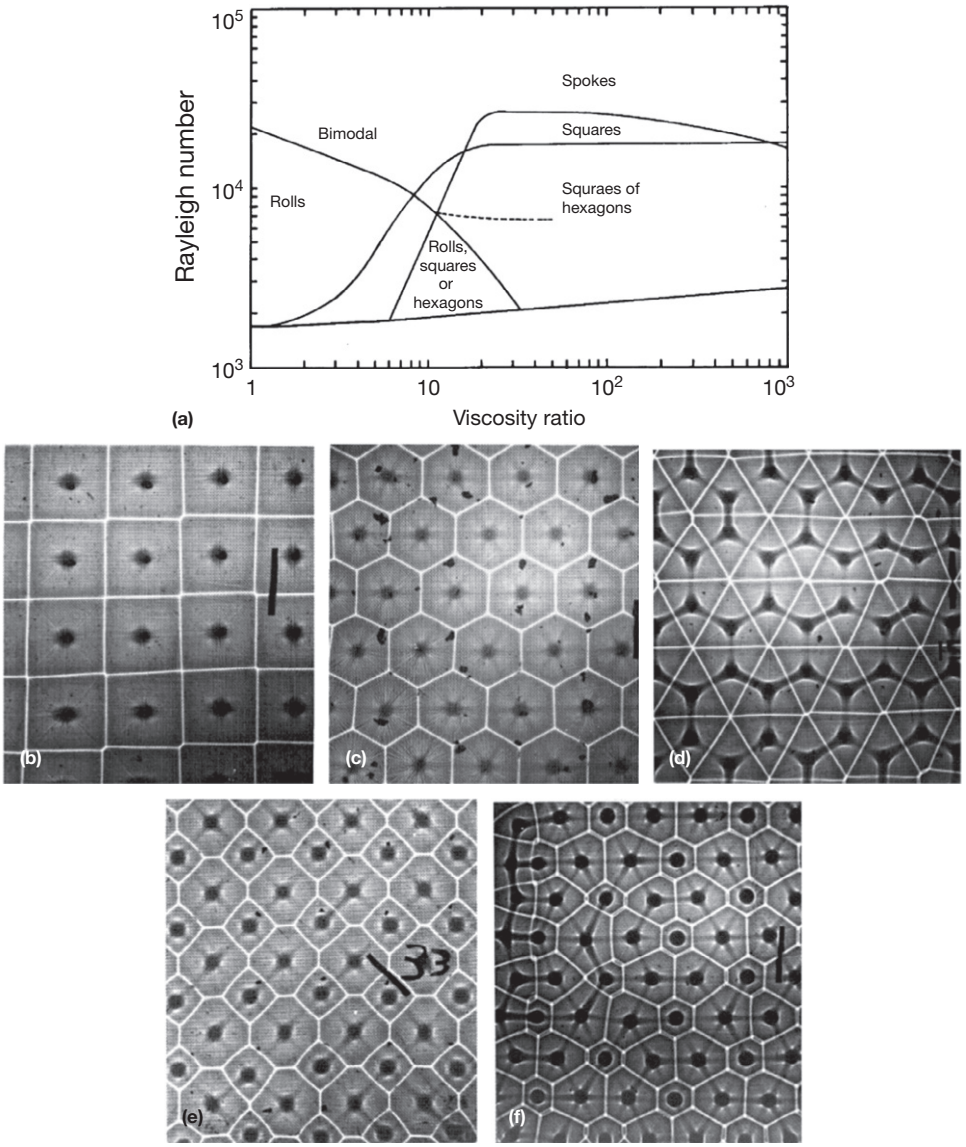


Figure 34 Convective patterns in fluids with strongly temperature-dependent viscosity (golden syrup) as a function of $Ra_{1/2}$ and γ. (a) Stability map for rolls, hexagons, and squares. There is a large amount of hysteresis where the realized platform depends on what pattern was originally induced, (b) squares, (c) hexagons, (d) triangles, (e, f) mosaic instabilities. Reproduced from White DB (1988) The planforms and onset of convection with a temperature-dependent viscosity fluid. *Journal of Fluid Mechanics* 191: 247–286.

7.03.6.1.3 Characteristics of the stagnant lid regime

Davaille and Jaupart (1993, 1994) studied transient high-*Ra* convection in a layer of golden syrup suddenly cooled from above. With this setup, the viscosity contrast across the cold TBL, γ_{TBL}, ranged between 2 and 10^6. For $\gamma_{TBL} > 100$, convection developed only in the lower part of the TBL, and the upper part remained stagnant (Figures 33 and 36). The Rayleigh number and all the scaling laws therefore used the viscosity of the well-mixed interior, since it would be meaningless to characterize the flow with a viscosity corresponding to a temperature well within the stagnant fluid. At the onset of convection, the viscosity contrast across the unstable region is about 3; in fully developed convection, it reaches a typical value of 10. The temperature difference ΔT_{eff} across the unstable part of the lithosphere depends only on the interior temperature T_m and on a temperature scale controlled by the rheology ΔT_c (e.g., Davaille and Jaupart, 1993; Morris, 1982):

$$\Delta T_c(T_m) = \frac{\eta}{\frac{\partial \eta}{\partial T}}(T_m) \qquad [12]$$

Laboratory results then give $\Delta T_{eff} = 2.24 \Delta T_c$. A necessary condition for the existence of a stagnant lid is that the applied temperature difference exceeds a threshold value equal to ΔT_{eff}. All the dynamics of the system are then determined locally in the unstable part of the TBL. Following once again the phenomenological model of Howard (1964), the heat flux can therefore be written as

$$Q_s = Ck\left(\frac{\alpha g}{k\nu(T_m)}\right)^{1/3}\Delta T_m^{4/3} \qquad [13]$$

where the constant $C = 0.47 \pm 0.03$ is determined experimentally. This model also predicts well the convective onset time and the amplitude of the temperature fluctuations (Davaille and Jaupart, 1993). Subsequent numerical and experimental work also recovered the functional forms of eqns [12] and [13], with $\Delta T_{eff} = 2.2–2.6 \times \Delta T_c$ and $C = 0.4–0.9$ (e.g., Androvandi et al., 2011; Dumoulin et al., 2001; Gonnermann et al., 2004; Grasset and Parmentier, 1998; Manga et al., 2001; Solomatov and Moresi, 2000; Trompert and Hansen, 1998a). In the case of steady Rayleigh–Bénard convection at high Ra, the temperature difference in the hot TBL has been found to scale as $1.1–1.3 \times \Delta T_c$ (Manga et al., 2001; Solomatov and Moresi, 2000), which together with eqn [11] allows one to predict the temperature of the well-mixed interior (Manga et al., 2001; Solomatov, 1995). Moreover, since the cold and hot unstable TBLs have different characteristics, plumes are released from these layers with different frequencies, and cold, more viscous downwellings carry a greater temperature anomaly than their hot counterparts. Measurements show that whereas there is a single frequency for cold plume formation, hot plumes form with multiple frequencies and in particular may be triggered by cold sinking plumes (Schaeffer and Manga, 2001).

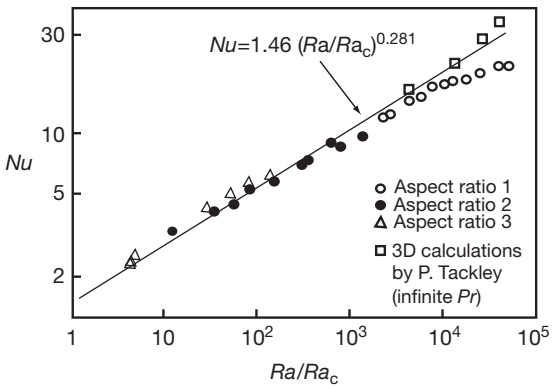


Figure 35 Nusselt number as a function of Rayleigh number. Reproduced from Manga M and Weeraratne D (1999) Experimental study of non-Boussinesq Rayleigh–Bénard convection at high Rayleigh and Prandtl numbers. *Physics of Fluids* 10: 2969–2976.

7.03.6.2 Stagnant Lid Regime and Lithosphere Cooling

The study of the stagnant lid regime is relevant for the dynamics of plate cooling. The oceanic lithosphere thickens at it moves away from a mid-ocean ridge (Parmentier, **Chapter 7.08**; **Chapter 6.05**). It has been proposed that cold thermal instabilities develop in its lower part. Such 'small-scale convection' (SSC) was first invoked to explain the flattening of heat flux and bathymetry of the oceans at old ages (Parsons and McKenzie, 1978; Parsons and Sclater, 1977) and later several other phenomena occurring on different time and length scales, such as small-scale (150–500 km) geoid anomalies in the central Pacific (e.g., Baudry and Kroenke, 1991; Haxby and Weissel, 1986) and central Indian (Cazenave et al., 1987) Oceans, differences in subsidence rates (Buck, 1987; Eberle and Forsyth, 1995; Fleitout and Yuen, 1984), ridge segmentation (e.g., Rouzo et al., 1995; Sparks and Parmentier, 1993), delamination of the lithosphere under hot spots (Dubuffet et al., 2000; Moore et al., 1998; Sleep, 1994), and patterns of anisotropy beneath the Pacific Ocean (Davaille and Jaupart, 1994; Ekstrm and Dziewonski, 1998; Montagner and Tanimoto, 1990; Nishimura and Forsyth, 1988). Davaille and Jaupart (1994) applied the scaling laws derived from their experimental results, using for the mantle a Newtonian creep law with activation enthalpy H:

$$\eta(T) = \eta_r \exp\left(\frac{H}{RT}\right) \qquad [14]$$

For H varying between 150 and 500 kJ mol^{-1} (the value depends on the creep mechanism and the water content; see Kolhstedt, volume 2), ΔT_{eff} ranges between 310 and 90 °C according to eqns [11] and [13]. Since small-scale instabilities develop when the local Rayleigh number based on the

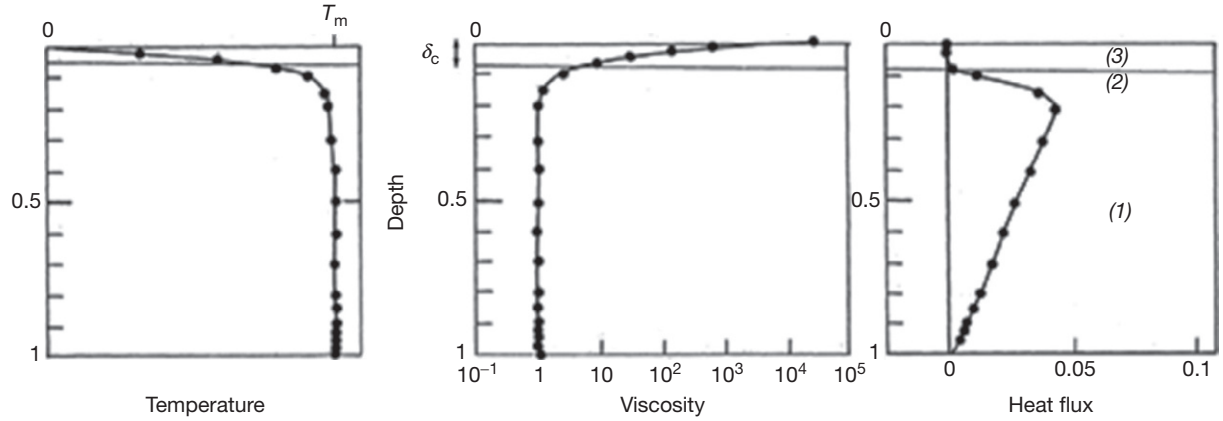


Figure 36 Structure of a layer of golden syrup cooled from above ($Ra = 2.5 \times 10^7$; $\gamma = 10^6$). (a) Horizontally averaged temperature profile. (b) Corresponding viscosity profile. (c) Convective heat flux calculated from the temperature time series. Its value is 0 is the topmost viscous part of the fluid layer (stagnant lid). Reproduced from Davaille A and Jaupart C (1993) Transient high-Rayleigh number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141166.

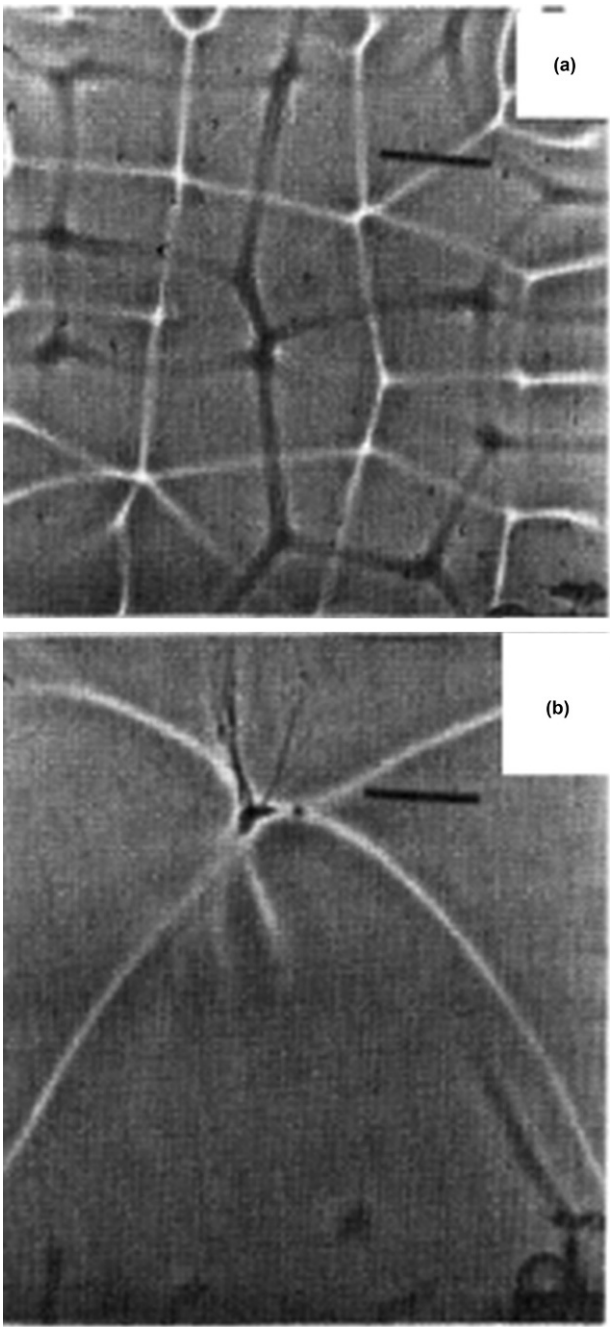


Figure 37 Shadowgraphs of the convection planforms found for the case of rigid boundaries (a) and for the case of a stress-free upper boundary (b). Reproduced from Weinstein SA, Christensen UR (1991) Convection planforms in a fluid with a temperature-dependent viscosity beneath a stress-free upper boundary. *Geophysical Research Letters* 18: 2035–2038.

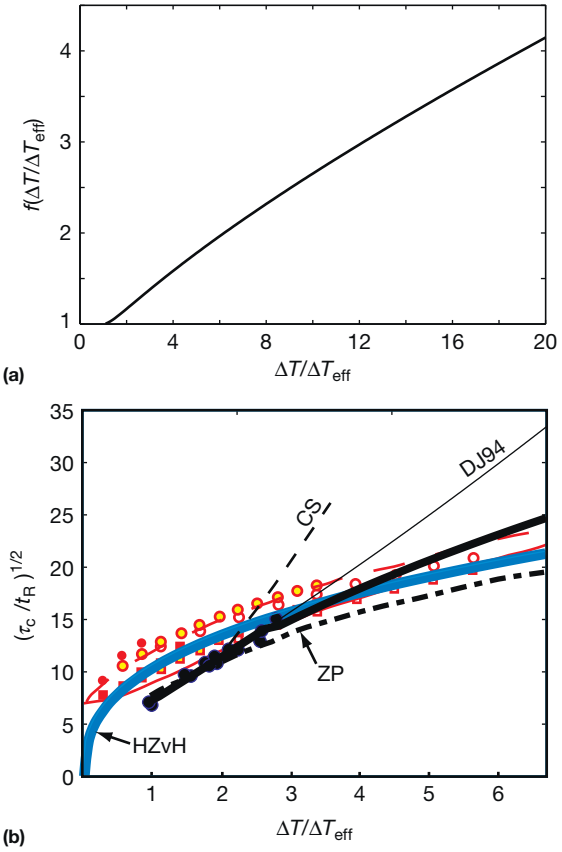


Figure 38 (a) Function $f(x)$ described in the text (eqn [7]) as a function of $\Delta T/\Delta T_{\text{eff}}$. (b) Square root of the dimensionless small-scale convection (SSC) onset time τ_c/t_R($t_R = k^{-1}(\alpha g \Delta T/k\nu)^{-2/3}$) as a function of $\Delta T/\Delta T_{\text{eff}}$. The black disks represent the laboratory data of Davaille and Jaupart (1994). The black thick solid line was calculated with eqn [7] and the thin line labeled 'DJ94' with the $f \sim 1$ approximation. The thin dashed black line labeled 'CS' shows Choblet and Sotin (2000) calculations, the blue thick line labeled 'HZvH,' Huang et al.'s (2003), and the black dash-dot line labeled 'ZP,' Zaranek and Parmentier's (2004). In red, data and scaling laws by Korenaga and Jordan (2003). Squares used an exponential viscosity law, and circles an Arrhenius law.

characteristics of the unstable part of the lithosphere (temperature, viscosity, and thickness) exceeds a critical value, the onset of convection depends only on the thermal structure of the lithosphere and on the rheology of mantle material. Following Davaille and Jaupart (1994), the onset time can be written as

$$\tau_c = \frac{b_1}{k}\left\{\frac{\alpha g \Delta T_{\text{eff}}}{k \nu_m}\right\}^{-2/3}\left[1 + \frac{b_2}{f\left(\mathrm{D}\frac{\Delta T}{\Delta T_{\text{eff}}}\right)}\left(\frac{\Delta T}{\Delta T_{\text{eff}}} - 1\right)\right]^2 \quad [15]$$

where $\nu_m = \eta(T_m)/\rho$ is the viscosity of the asthenosphere at the mantle temperature T_m, $b_1 = 51.84$ and $b_2 = 0.3013$ are two constants determined from the laboratory experiments, and ΔT is the temperature drop across the lithosphere. The function f depends on the cooling model, and for a half-space, conductive cooling is given in **Figure 38(a)**. For a thin stagnant lid (or $\Delta T/\Delta T_{\text{eff}} < 3$), $f \sim 1$. This approximation applies well to laboratory experiments and was therefore adopted by Davaille and Jaupart (1994) (thin line labeled DJ94 on **Figure 38(b)**). However, as pointed out by subsequent studies (Dumoulin et al., 2001; Zaranek and Parmentier, 2004), $\Delta T/\Delta T_{\text{eff}} > 4$ for the mantle lithosphere and the approximation $f \sim 1$ leads to overestimate the SSC onset time (**Figure 38(b)**) compared to the prediction of eqn [15] without the approximation (thick black line on **Figure 38(b)**). In the last years, several numerical studies,

using either Arrhenius or exponential viscosity laws, have allowed to extend the range of $\Delta T/\Delta T_{\text{eff}}$ studied (Choblet and Sotin, 2000; Dumoulin et al., 2001; Huang et al., 2003; Korenaga and Jordan, 2003; Zaranek and Parmentier, 2004) and to develop new scaling laws. The discrepancies between the different data sets reveal the relative importance of the onset time measurement, the type of viscosity law (10–20% of change only), and the type of perturbation (initial time/at all times and numerical noise/finite-size perturbation) introduced numerically that will grow toward convective instabilities. The latter has the strongest influence (Zaranek and Parmentier, 2004): the longest onset times are obtained when the numerical perturbations initially introduced are smallest (Choblet and Sotin, 2000). The trend of the laboratory experiments seems to compare best with numerical experiments, which introduce thermal noise around $10^{-3} \times \Delta T$ at all times (**Figure 38(b)**). Then, in the parameter range relevant for the Earth's mantle ($\Delta T/\Delta T_{\text{eff}} \geq 4$), numerical studies agree with eqn [15] within 25%. So, if the magnitude of the temperature perturbations in the lithospheric mantle is similar to the laboratory one, an SSC onset time between 10 and 100 My is predicted, depending on the asthenospheric mantle temperature.

7.03.6.3 Plate Tectonics in the Laboratory?

7.03.6.3.1 Free slip boundaries

Mantle convection is characterized by mobile plates on its top surface, which have no chance to be recovered in experiments with a rigid top boundary since the latter imposes a zero fluid velocity on the surface. Moreover, 2-D numerical simulations had shown that the heat transport strongly depends on the mobility of the surface boundary layer (Christensen, 1985), which is controlled by the viscosity contrast. Giannandrea and Christensen (1993) and Weinstein and Christensen (1991) therefore carried out experiments in syrups in a large aspect ratio tank (**Figure 37**) to study the effect of a stress-free boundary (**Figure 5(b)**). They observed two regimes. For $\gamma \leq 1000$, the surface layer is mobile. At $Ra = 10^5$, the morphology of downwellings changes to a dendritic network of descending sheets, with a wavelength more than three times that of the spoke pattern (**Figure 37**). The existence of this 'whole-layer' or 'sluggish lid' mode with an increased wavelength is in agreement with earlier predictions (Jaupart and Parsons, 1985; Stengel et al., 1982) and 3-D numerical calculations (Ogawa et al., 1991). The Nusselt number drops by 20% for viscosity contrasts between 50 and 5000. For $\gamma \geq 1000$, a stagnant lid forms on the top of an actively convecting region; and both the Nusselt number and the size of the convection cells are nearly identical for both rigid and free boundaries.

7.03.6.3.2 Several scales of convection

On Earth, seismic tomography studies show that the subduction zones distribution – and therefore the cold plates distribution – delimits two large cells in the mantle (the Pacific and the Indo-Atlantic 'boxes'), each cell containing several hot instabilities (plumes). According to **Figure 33** (pink area), the existence of several scales of convection suggests that the Earth's mantle is in the 'mobile lid' regime (e.g., Androvandi et al., 2011; Solomatov, 1995). Given the mantle physical characteristics, this would imply that the effective viscosity of the lithosphere is not more than ~30–3000 times more viscous than the bulk mantle (Androvandi et al., 2011). Such values are consistent with estimates using bathymetry and geoid constraints (Billen et al., 2003), regional models of subduction (Funiciello et al., 2008; Ribe, 2010; Schellart, 2008; Schellart et al., 2007; Stegman et al., 2010; see **Section 7.03.11**), or slab bending (Wu et al., 2008).

7.03.6.3.3 'True' plate tectonics

However, according to the regime diagram of **Figure 33**, with an actual viscosity contrast across the lithosphere $\geq 10^7$, the Earth should be in the stagnant lid regime, that is, a one-plate planet like Mars. So, if temperature-dependent viscosity is clearly a key ingredient for plate formation, this ingredient alone is not sufficient to generate plate tectonics convection. To make plate tectonics work, weakening and shear localization are also required (see Bercovici, 2003; Bercovici et al., 2000; Ricard, **Chapter 7.02**; Bercovici et al., **Chapter 7.07**). Pseudoplastic yielding is relatively easy to implement in numerical simulations, whereby plate boundaries develop instantaneously when the local stresses exceed a certain yield stress (Grigné et al., 2005; Moresi and Solomatov, 1998; Stein et al., 2004; Tackley, 2000a,b; Trompert and Hansen, 1998b). However, this method requires a yield stress much lower than what is known for mantle material. Gurnis et al. (2000) also pointed out the importance of lithosphere 'memory': the lithosphere can support dormant weak zones (faults or rifts) over long time periods, but those weak zones can be preferentially reactivated to become new plate boundaries. A combination of damage (e.g., Bercovici et al., 2001a,b) and Zener pinning at grains boundary (e.g., Bercovici and Ricard, 2012) allows both the development and the slow cicatrization over time of weak zones. It gives a coherent framework to mantle deformation from the grain to the lithospheric scale (Bercovici and Ricard, 2013; Bercovici et al., this volume).

For years, experimentalists have been looking for an analogue fluid presenting the 'right' kind of rheology to allow self-consistent, true plate tectonics. The temperature dependence of the rheology of wax has proved to be too abrupt across its liquid to solid transition to generate plate tectonics. Such fluids are more adequate to model lava lake dynamics (e.g., Karlstrom and Manga, 2006). Gelatins are extensively used to study crustal dynamics but their strong elastic and brittle behavior precludes a good similarity to mantle flow (Di Giuseppe et al., 2009). Recent experiments using aqueous colloidal dispersions of nanoparticles seem more promising. The rheology of these colloidal solutions depends strongly on the solid particle fraction ϕ_p and on the chemical activity a_χ, being Newtonian at low ϕ_p and low a_χ and presenting yield stress, elasticity, and brittle properties when ϕ_p or a_χ increase (Di Giuseppe et al., 2012). So by heating these fluids from below, as well as drying them (as an analogue to cooling) from above, several regimes can be observed whereas a viscoelastoplastic skin forms at the surface and either stays there (convection under a stagnant lid) or breaks and sinks episodically in the fluid (episodic subduction, **Figure 39**, Di Giuseppe and Davaille, 2010), or presents the continuous motion and renewal of surface plates (Davaille et al., 2011). The latter regime is close to the continuous plate tectonics observed on Earth, and subduction is always asymmetrical.

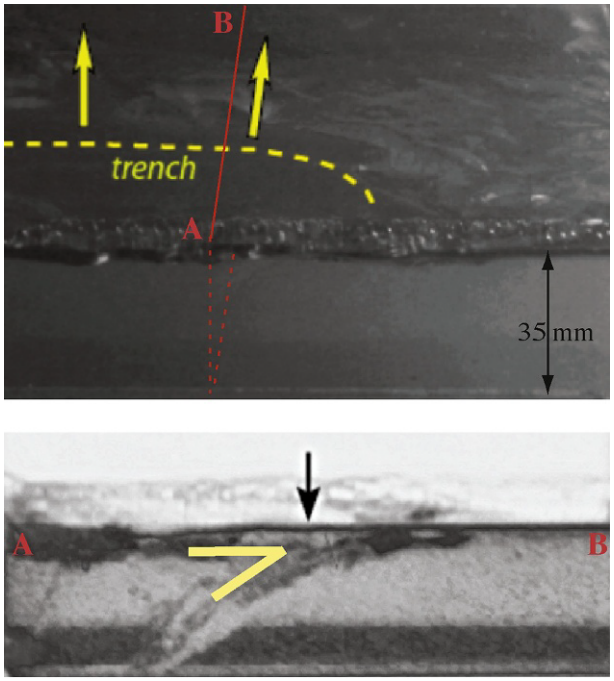


Figure 39 Spontaneous asymmetrical subduction in a layer of colloids dried from above at constant temperature and humidity. (a) View from above. The 'trench' is outlined in yellow and the direction of trench rollback is indicated by the yellow arrows. (b) View from the side. The contrast has been enhanced to better see the subducting plate. The trench position is marked with a black arrow.

7.03.7 Complications: Internal Heating, Continents, and Moving Plates

7.03.7.1 Internal Heating

We have so far focused on the dynamics of fluid layers uniformly heated from below and cooled from above. However, the Earth's mantle is not only heated along the CMB by the hotter molten iron core but also partly heated from within by radiogenic decay of uranium, thorium, and potassium distributed throughout the mantle (see Jaupart et al., **Chapter 7.06**). Thermal convection with both internal and bottom heating is characterized by two dimensionless numbers: Ra for the bottom heating given by eqn [1] and an equivalent Rayleigh number for the purely internally heated case, defined as

$$Ra_H = \frac{\alpha g H^5 Q_H}{\nu k \kappa} \qquad [16]$$

where Q_H is the volumetric rate of internal heat generation.

7.03.7.1.1 Purely internally heated fluid

Internal heat sources are usually simulated by the ohmic dissipation of an electric current passed through an electrolyte fluid layer. Electrodes are placed either on each of the horizontal boundaries (e.g., Carrigan, 1982) or on two of the sidewalls of the chamber (e.g., Kulacki and Goldstein, 1972). Polarization effects in the fluid are eliminated by applying an alternating current. In all the experiments using this technique, the bottom of the tank is insulated (zero heat flux) and the top of the tank is cooled by a thermostated bath. One therefore expects only one TBL in the system, just below the cold plate.

Most of the experiments that have been performed used aqueous solutions. Tritton and Zarraga (1967) and Schwiderski and Schwab (1971) visualized the planform of convection from marginal stability up to the transition to turbulence. Quasi-steady planforms were observed with downflow in the centers of the cells for $Ra_H \leq 80.Ra_c$. Recent studies (Takahashi et al., 2010; Tasaka et al., 2005) further showed that a spoke-like structure was stable in ideal infinite lateral conditions but that a double-cell structure was formed if finite lateral boundaries were imposed. For larger Ra, cells broke up and turbulent motions appeared. Kulacki and Goldstein (1972) extended the range up to $675.Ra_c$, well into the turbulent regime. Kulacki and Nagle (1975) focused on heat flow measurements as a function of Ra_H. For the turbulent regime ($1.5 \times 10^5 \leq Ra_H \leq 2.5 \times 10^9$), their data are consistent with a power law $Nu \sim 0.305 \times Ra^{0.239}$. As for the bottom-heated case, transitions in the $Nu = f(Ra_H)$ curves were observed at each change of the convective pattern. Thermal fluctuations were also found to be maximum just outside the upper cold TBL, and no bottom hot TBL was observed.

Carrigan (1982) used glycerine ($Pr \sim 3000$) for $10^5 \leq Ra_H \leq 10^6$. Cold spoke patterns were observed up to $Ra_H = 108.Ra_c$, while a superposition of a large-scale circulation and cold individual plumes developed for larger Ra. More recently, the thermal and temperature fields for Ra_H up to 10^7 ($Pr \sim 2800$) were obtained in a Natrosol aqueous solution with microwave-generated internal heating (Limare et al., 2013).

7.03.7.1.2 Bottom-heated and internally heated fluid

Following Krishnamurti (1968a), Weinstein and Olson (1990) demonstrated the equivalence between thermal convection in a plane layer of fluid with uniformly distributed heat sources between boundaries with constant temperatures and thermal convection in a plane layer of fluid with boundary temperatures that are spatially uniform but vary in time at a rate dT_{top}/dt. The effective volumetric internal heat generation is then

$$Q_H = \rho C_P \frac{dT_{top}}{dt} \qquad [17]$$

This technique is simpler and more flexible to use than the electrolytic or the microwave ones, since it can be used with any homogeneous fluid and allows the fluid to be heated from below as well as internally. Krishnamurti (1968b) performed experiments with silicone oils ($Pr \sim 200$) near the onset of convection. Using high-Prandtl-number oils ($Pr \sim 9 \times 10^3$), Weinstein and Olson (1990) studied the evolution of the convective pattern and its time dependence as the proportion of internal heating increases (**Figure 40**). At $Ra = 1.5 \times 10^5$, in the absence of internal heat generation, a spoke pattern is observed, in which connected spokes of ascending and descending flows are on average of equal strength. With internal heat generation, the buoyancy becomes concentrated into the cold downwellings, and the amount of time dependence increases. Hot active instabilities will reappear if there exists a bump on the hot bottom boundary with a height comparable to (or greater than) the TBL thickness (Namiki and Kurita, 2001). In this case, hot plumes will be anchored on the

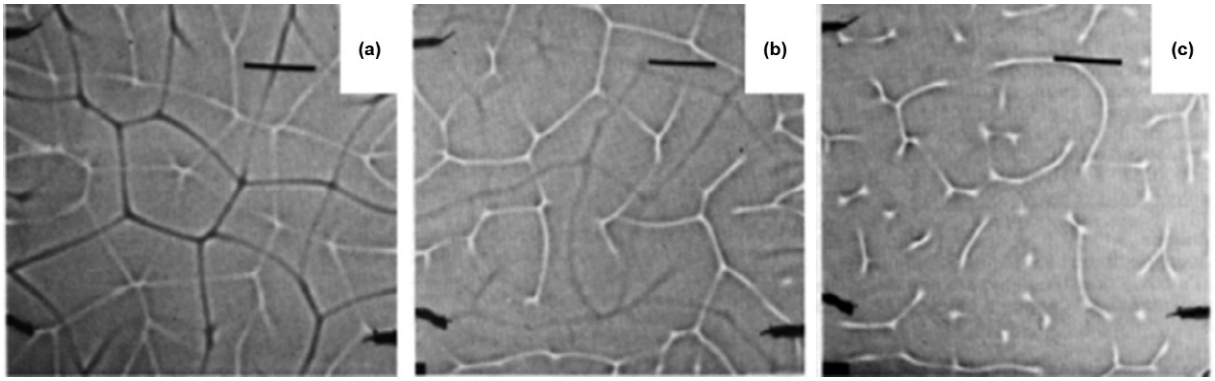


Figure 40 Shadowgraphs of the planform of thermal convection with internal heat generation for $Ra = 1.5 \times 10^5$ (spoke pattern) (a) $Ra_H/Ra = 0$; light and dark lines have the same intensity as cold and hot currents have the same strength. (b) $Ra_H/Ra = 6$; the dark lines are fading as the downwellings become weaker. (c) $Ra_H/Ra = 18$; no hot upwellings are seen anymore. Reproduced from Weinstein SA and Olson P (1990) Planforms in thermal convection with internal heat sources at large Rayleigh and Prandtl numbers. *Geophysical Research Letters* 17: 239–242.

bump. So although an internally heated mantle favors cold downwellings, hot plumes can still develop, provided that the bottom of the mantle, or the D″ layer, has some topography.

7.03.7.2 Inhomogeneous Cooling/Continents

Continental lithosphere is thicker than its oceanic counterpart and does not subduct. Its crust is also richer in radioactive elements. Moreover, heat flow beneath continents is significantly lower than under oceans (e.g., Jaupart et al., 1998), which tends to show that continents are more insulating. Hence, continents induce a lateral heterogeneity in both the mechanical and thermal conditions at the upper boundary of the mantle. The experimental study of Rossby (1965) showed that any lateral heterogeneity in cooling or heating will generate a horizontal large-scale flow. Elder (1967) was probably the first to suggest that the continents, by acting as an insulating lid, could generate a new class of lateral motions. Later studies have therefore focused on characterizing and quantifying the influence of these lateral heterogeneities on mantle thermal convection.

Whitehead (1972, 1976) investigated both theoretically and experimentally the conditions under which a heater mounted on a mobile float could become unstable to drifting motion. For cellular convection, the drift was found of the same magnitude as the characteristic convective velocity.

Guillou and Jaupart (1995) investigated the thermal effect of a fixed continent on the convective pattern and heat flow, using silicone oils and golden syrup for Ra ranging between critical and 10^6. The continent was modeled as an insulating material embedded into a cold copper plate (**Figure 41**), and its size and shape were varied. They observed the generation of deep hot instabilities localized under the continent and the generation of cold instabilities outside or on its edges (**Figure 41**). The presence of a continent therefore shields part of the underlying mantle from subduction (Gurnis, 1988) and also induces a large-scale convection pattern

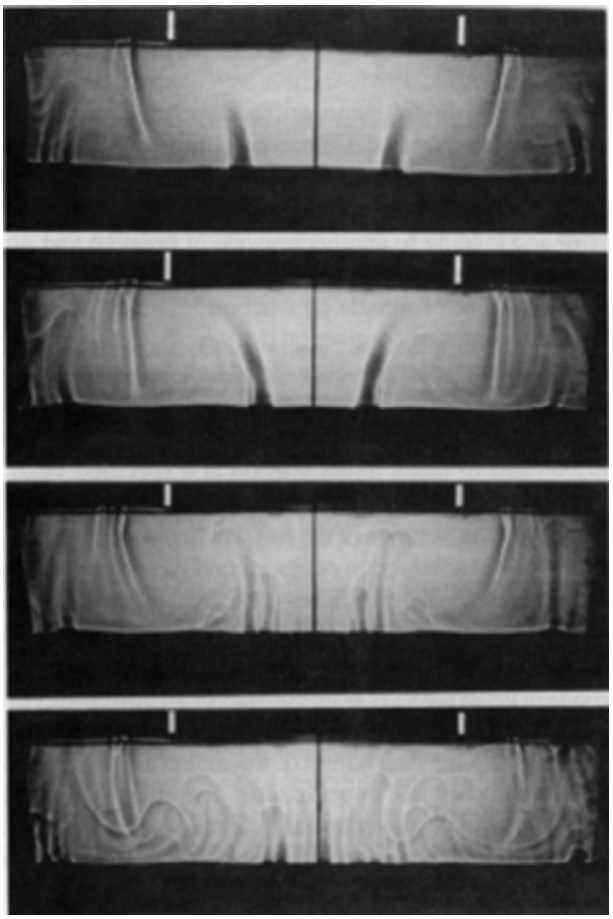


Figure 41 Insulating continents from Guillou and Jaupart (1995). Markers along the upper boundary indicate the extent of insulating material representing continental lithosphere. Time progresses from top to bottom. Note upwelling plumes beneath the interior and descending sheets near the margins of the insulator.

oceanward over distances up to three times the continent's size (Guillou and Jaupart, 1995). Lenardic and Jellinek (Jellinek and Lenardic, 2009; Lenardic et al., 2005, 2011) extended the study to Ra up to 2×10^8 and γ up to 10^5 and investigated the influence of partial insulation on mantle circulation and heat extraction. Partial insulation leads to increased internal mantle temperature and decreased viscosity, either under the supercontinent or in the whole mantle, depending on the distribution and rheology of the cold downwellings and on the ratio γ of continental to oceanic surface so that global mantle heat flow goes through a maximum for $\gamma \sim 0.4$–0.6 (Jellinek and Lenardic, 2009). Therefore, the breakup of a supercontinent to a number of small continents should generate sizable bulk mantle temperature and heat flux variations (Lenardic et al., 2011). Wenzel et al. (2004) used the same kind of setup with two-layer convection. They suggested that, on Mars, the combination of crustal dichotomy (the Martian crust is much thicker, therefore insulating, in the southern hemisphere than in the northern hemisphere) and two-layer convection would lead to the formation of a large-scale upwelling under the southern highlands that appeared early and endured for billions of years.

Zhang and Libchaber (2000) and Zhong and Zhang (2005) replaced the rigid cold boundary by a free surface cooled by air, which allowed the insulating raft to move freely in response to the convection (**Figure 42**). The experimental fluid was water, with Rayleigh numbers ranging from 10^7 to 4×10^8. Like Guillou and Jaupart (1995), they observed that the insulating raft distorts the local heat flux and induces a coherent flow over the experimental cell, which in turn moves the plate. In their confined geometry, the plate can be driven to a periodic motion even under the action of chaotic convection. The period of oscillation depends on the coverage ratio and on the Rayleigh number (**Figure 42**).

Nataf et al. (1981) investigated the influence of the sinking cold oceanic lithosphere on the neighboring subcontinental convection and on the generation of instabilities in the horizontal boundary layers by performing experiments with cooling from both the top plate and one of the sides. The system dynamics are then characterized by two Rayleigh numbers: Ra based on the vertical temperature difference and a lateral Rayleigh number based on the lateral temperature difference (Ra_{lat}). The lateral cooling induces a large roll with axis parallel to the cold wall and descending flow next to it. At constant Ra, the width of the large roll is proportional to $Ra_{\text{lat}}^{1/2}$, while for constant Ra_{lat}/Ra, it scales as $\sim Ra^{1/6}$. In the upper mantle, such large-scale circulation could become 5 times wider than deep. The experiments also show that hot instabilities cannot develop directly beneath the cold downwelling flow, because the spreading of the latter impedes the growth of the hot TBL.

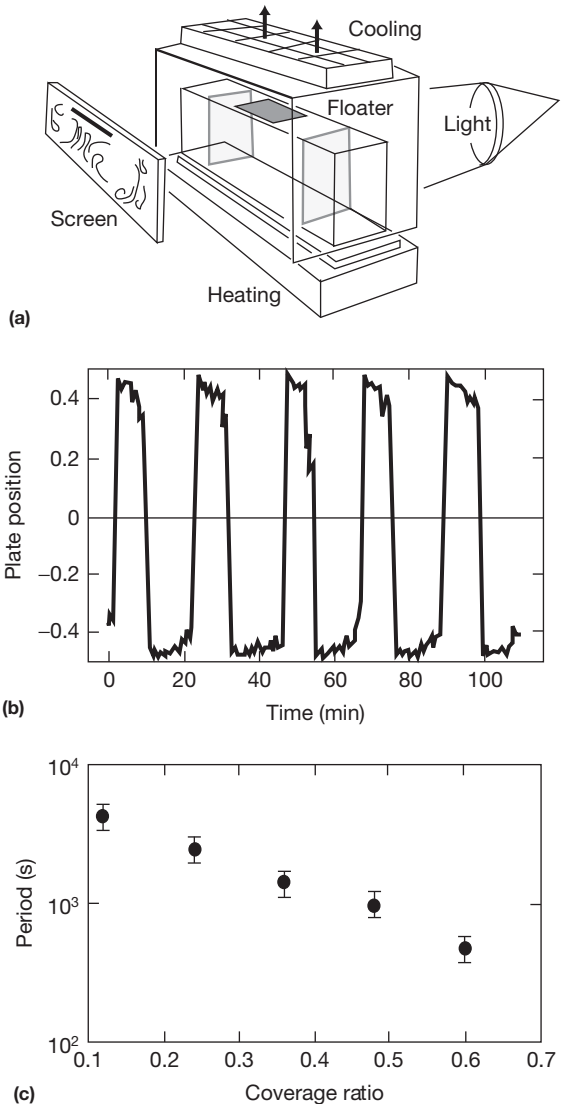


Figure 42 Mobile insulating continent. (a) Experimental setup. (b) Continent position as a function of time. (c) Continental drift periodicity as a function of the coverage ratio (length of the floater divided by tank length). Reproduced from Zhang J and Libchaber A (2000) Periodic boundary motion in thermal turbulence. *Physical Review Letters* 84: 4361–4364.

7.03.7.3 Moving Top Boundary

Any large-scale circulation strongly influences the generation of convective features of smaller extent. The large-scale motion of plates is therefore expected to influence all motions within the mantle, especially SSC under the cooling lithosphere, and the distribution of hot spots.

Following the theoretical study of Richter (1973), Richter and Parsons (1975) studied the influence of a moving top boundary on the convective pattern for Ra between 3.3×10^3 (rolls) and 1.4×10^5 (spokes). A mylar sheet was introduced just below the top cold boundary and pulled at constant velocity U (**Figure 5(a)**). Besides Ra, another system now controls the system dynamics, the Péclet number $Pe = UH/\nu$, which compares the heat transport by advection and by diffusion. It ranges between 30 and 720 in the experiments. Initially, the top boundary was at rest until the convective pattern becomes quasi-steady. For the lower Ra, rolls orthogonal to the spreading direction were initiated using the technique of Chen and Whitehead (1968). Then, the velocity of the mylar was set to a finite value and the convective pattern followed by shadowgraph. Over the whole Ra range, the pattern was observed to evolve toward rolls parallel to the mylar velocity. Scaling laws show that it would probably take 70 My for the geometry of

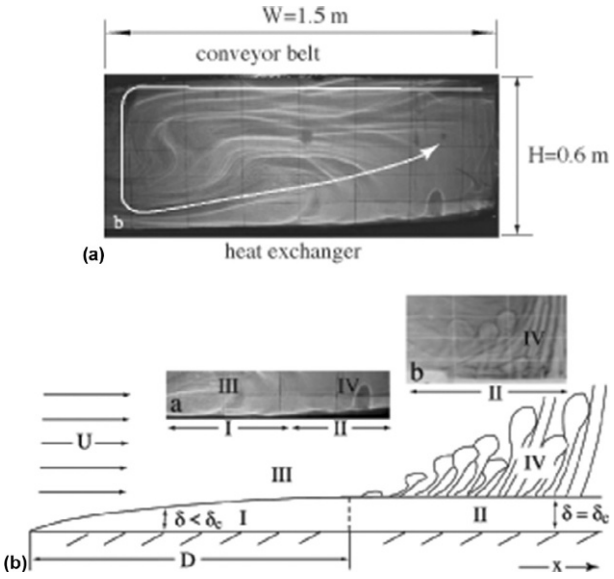


Figure 43 (a) Experimental setup and shadowgraph at $Ra \sim 5 \times 10^6$, $Pe \sim 10^4$, and $\gamma \sim 100$. Note the increasing thickness of the TBL in the downstream direction. (b) Schematic diagram depicting the four convecting regions. The TBL is not thick enough in region I to develop plumes in III, while TBL instabilities develop in regions II and IV. Reproduced from Gonnermann HM, Jellinek AM, Richards MA, and Manga M (2004) Modulation of mantle plumes and heat flow at the core mantle boundary by plate-scale flow: Results from laboratory experiments. *Earth and Planetary Science Letters* 226: 5367.

convection to reorganize into rolls parallel to the spreading direction over the whole thickness of the upper mantle.

Experiments of Kincaid et al. (1996), Jellinek et al. (2002, 2003), and Gonnermann et al. (2004) studied how a plume-dominated free convective flow interacts with a large-scale passive flow driven by plate motions. Kincaid et al. (1996) focused on the interaction of mid-ocean ridges and convection in the upper mantle. The three other papers extended this study to the whole mantle, higher *Ra*, and fluids with strongly temperature-dependent viscosity. Jellinek et al. (2002, 2003) performed laboratory experiments using corn syrup heated from below and driven at its surface by a ridge-like velocity boundary condition (**Figure 43(a)**). They documented three different flow regimes, depending on the ratio V of the spreading rate to the typical plume rise velocity and the ratio of the interior viscosity to the viscosity of the hottest fluid. When $V \ll 1$ and $\gamma \sim 1$, the plume distribution is random and unaffected by the large-scale flow. The opposite extreme is represented by $V > 10$ and $\gamma > 100$, where plume formation is suppressed entirely and the large-scale flow carries all the heat flux. Interaction of the two flows occurs for intermediate values of V; established plume channels are now advected along the bottom boundary and focused toward an upwelling beneath the ridge. The strong localization of the upwelling beneath the ridge is here due to the small aspect ratio (=2.5) of the experimental tank and the fixed position of the ridge. However, in a larger aspect ratio box like the mantle, upwellings will still developed mainly away from subducting plates. The basal heat flux and the TBL thickness depend on *Pe* and γ. In particular, since cold, more viscous fluid is rapidly transported from the top of the layer, the temperature contrast across the lower TBL increases, generating hot plumes with a bigger head to tail ratio than without plate circulation (Jellinek et al., 2002; Nataf, 1991). Running more detailed experiments, Gonnermann et al. (2004) proposed a conceptual framework (**Figure 43(b)**) whereby the tank bottom can be divided in four regions. The large-scale flow impedes the growth of the hot TBL where the cold and more viscous downwelling spreads (Nataf et al., 1981) and therefore the hot TBL is not thick enough in region I to develop plumes in III. In region II, the TBL becomes thick enough for hot instabilities to develop in IV. Scalings were derived for the lateral variations of the bottom heat flux and the spatial extent of plume suppression. Applied to the Earth's mantle, they suggest that ~30% of the core heat flux could be due to increased heat flux from plate-scale flow. Furthermore, the core heat flux is expected to vary laterally by at least a factor of 2. Last, these results explain well why hot spots seem confined to about a half of the surface of the Earth, away from ancient subducted slabs, and preferentially close to mid-ocean ridges (Weinstein and Olson, 1989). This behavior is consistent with what is observed in the mobile lid regime of thermal convection at high *Ra* and intermediate γ (Androvandi et al., 2011; **Figures 32(d)** and **33**).

7.03.8 Close-Up on Plumes: Plumes from a Point Source of Buoyancy

Since the mid-1970s, much of the best experimental work in mantle dynamics has been devoted to plumes from point sources of buoyancy. While plumes in the mantle probably develop primarily as instabilities of TBLs (e.g., Loper and Stacey, 1983; Parmentier et al., 1975), such instabilities are time-dependent and can be difficult to quantify (see **Section 7.03.5**). As a first step, therefore, it makes sense to investigate the simpler case of an isolated laminar starting plume rising from a point source of buoyancy whose strength is constant in time. Such plumes are easily studied and photographed in the laboratory and have given us some of the most beautiful images in the field of experimental geodynamics (e.g., Campbell and Griffiths, 1990; Griffiths, 1986a,b; Olson and Singer, 1985; Whitehead and Luther, 1975). Indeed, these images have been so influential that they now constitute a sort of standard model of a mantle plume as a large, bulbous cavity (the head) trailed by a narrow conduit (the tail) connecting it with its source (**Figure 7**). According to this model, the arrival of a plume head at the base of the lithosphere produces massive flood basalts, while the trailing conduit generates the subsequent volcanic track (Richards et al., 1989). The model successfully explains important features of several prominent hot spots, including the volume of the topographic swell (e.g., Davies, 1988; Olson, 1990; Sleep, 1990) and the volume ratio between flood basalts and island chain volcanism (Olson and Singer, 1985; Richards et al., 1989), as discussed in more detail by Ballmer et al. in **Chapter 7.10** of this volume.

7.03.8.1 Compositionally Buoyant Plumes

The early experiments on starting plumes in viscous fluids (Olson and Singer, 1985; Whitehead and Luther, 1975) were done by injecting compositionally buoyant fluid (typically corn

syrup) at a constant rate from a pipe (of radius r_p) at the bottom of a large reservoir of a second fluid. The great advantage of this experimental setup is its simplicity: experiments can be performed under ambient laboratory conditions, with no temperature control required. Whitehead and Luther (1975) investigated the evolution of starting plumes for short times after the beginning of injection, during which the buoyant fluid forms a quasi-spherical ball that grows until it is large enough to 'lift off' from the injector (Figure 7). They proposed that the ball lifts off when its buoyant ascent speed (as predicted by Stokes law; see Chapter 7.04) exceeds the rate of increase of its radius, which occurs at a critical time t_{sep} and radius a_{sep} given by

$$t_{sep} = \left(\frac{4\pi}{3Q}\right)^{1/4} \left(\frac{\nu_m}{g^*}\right)^{3/4} \quad [18]$$

$$a_{sep} = \left(\frac{3Q}{4\pi}\right)^{1/3} t_{sep}^{1/3} \quad [19]$$

where Q is the volumetric injection rate, ν_m is the viscosity of the ambient fluid, $g^* = g(\rho_m - \rho_p)/\rho = g\Delta\rho/\rho$ is the reduced gravitational acceleration, and ρ_m and ρ_p are the densities of the ambient and injected fluids, respectively. Both the time and radius of separation increase with increasing viscosity of the ambient fluid. For the large viscosities typical of the mantle, plume heads must be quite large (radius $\gg 100$ km) to separate from their source (Whitehead and Luther, 1975). After liftoff, the morphology and dynamics of starting plumes depend strongly on the viscosity ratio $\gamma = \nu_m/\nu_p$, where ν_p is the viscosity of the injected fluid (Olson and Singer, 1985; Whitehead and Luther, 1975). When the injected fluid is more viscous ($\gamma > 1$), the plume has roughly the shape of a cylinder (Figure 7(a)) whose length and radius increase with time. Olson and Singer (1985) called this a 'diapiric' plume. In the more geophysically realistic limit $\gamma \ll 1$, the ascending plume is a 'cavity plume' comprising a large head and a thin trailing conduit (Figure 7(b)). The rate of change of the volume V of the head is just the injected flux Q less the flux required to build the (lengthening) stem, or

$$\frac{dV}{dt} = Q - \pi R_c^2 W \quad [20]$$

where R_c is the radius of the conduit required to carry the flux Q by Poiseuille (cylindrical pipe) flow and W is the velocity of the head predicted by Stokes law. For long times, a steady state ($dV/dt = 0$) is achieved in which the head reaches a terminal velocity W_{term} and radius a_{term} (Whitehead and Luther, 1975):

$$W_{term} = \left(\frac{g^* Q}{8\pi\nu_p}\right)^{1/2} \quad [21]$$

$$a_{term} = \left(\frac{3W_{term}\nu_m}{g^*}\right)^{1/2} \quad [22]$$

More recent experiments have shown that under certain conditions, compositionally buoyant cavity plumes can entrain the denser ambient fluid through which they rise (Kumagai, 2002; Neavel and Johnson, 1991). Kumagai (2002) identified two distinct regimes: a 'vortex ring' regime for $\gamma < 10$ in which layers of entrained fluid form a scroll-like pattern inside the plume head and a 'chaotic stirring' regime for $\gamma \geq 100$ (see Section 7.03.10; Figure 57). Rogers and Morris (2009) extended the range of intrusion rate and studied the transition from laminar, buoyancy-driven plumes to momentum-driven jets. The key parameter is now the Richardson number $Ri = g^* r_p/W$. For $Ri > 1$, the plumes are buoyancy-driven and rise at velocity W_{term}, and their heads present the classical vortex ring. For $Ri < 1$, head dispersal is observed by a breakdown of the overturning motion in the head and a local Kelvin–Helmholtz instability on the exterior of the plume. Structures of the vortex ring type also occur in thermal starting plumes (Davaille et al., 2011; Griffiths and Campbell, 1990, Figure 7(c); Lohmann et al., 2009; Shlien, 1976; Shlien and Boxman, 1981; Tanny and Shlien, 1985).

The stem of a chemically buoyant plume can also support solitary waves, waves of large amplitude that propagate upward without change in shape. Their existence was first demonstrated experimentally by Olson and Christensen (1986) and Scott et al. (1987), who generated the waves by increasing impulsively the volume flux Q injected at the bottom of a vertical conduit surrounded by a fluid with a much greater viscosity. Whitehead and Helfrich (1988) subsequently showed that the flow within such waves exhibits recirculation along closed streamlines when viewed in a frame traveling with the wave and suggested that this might provide a mechanism for transporting deep material rapidly to the surface with little contamination. Subsequent numerical experiments (Schubert et al., 1989) showed that thermal plume stems could also support solitary waves, and Laudenbach and Christensen (2001) (Figure 44) performed an experimental study of solitary waves of hot fluid in a vertical conduit. Solitary waves on the stems of thermal plumes have been invoked to explain episodic magma production at weak hot spots and surges of activity at stronger hot spots (Laudenbach and Christensen, 2001) and the formation of the V-shaped ridges on the Reykjanes Ridge south of Iceland (Albers and Christensen, 2001; Ito, 2001).

7.03.8.2 Thermal Plumes

Because the buoyancy force that causes plumes to rise in the Earth's mantle is to a large degree thermal in origin, numerous laboratory investigations of thermal plumes in viscous fluids have been carried out since the mid-1980s. The first experiments on laminar starting plumes (Shlien, 1976; Shlien and Thompson, 1975) were not motivated by geophysical applications and were performed using water and a narrow electrode as a heater. Experiments in this configuration typically used interferometry and particle tracking to recover the temperature (Figure 13(c)) and velocity fields (e.g., Chay and Shlien, 1986; Shlien and Boxman, 1979; Tanny and Shlien, 1985). By contrast, investigations of plumes in more viscous fluids have until now mostly focused on more global features such as the volume and ascent rate of the plume head. Because the viscosity of many experimental fluids decreases strongly with increasing temperature, hot thermal starting plumes generally are cavity plumes comprising a large head and a narrow stem (e.g., Figure 45(a) and 45(e)). In some experiments, however, the heat source is suddenly turned off to produce an isolated head ('thermal') that rises without a trailing conduit. The discussion in the succeeding text begins with this

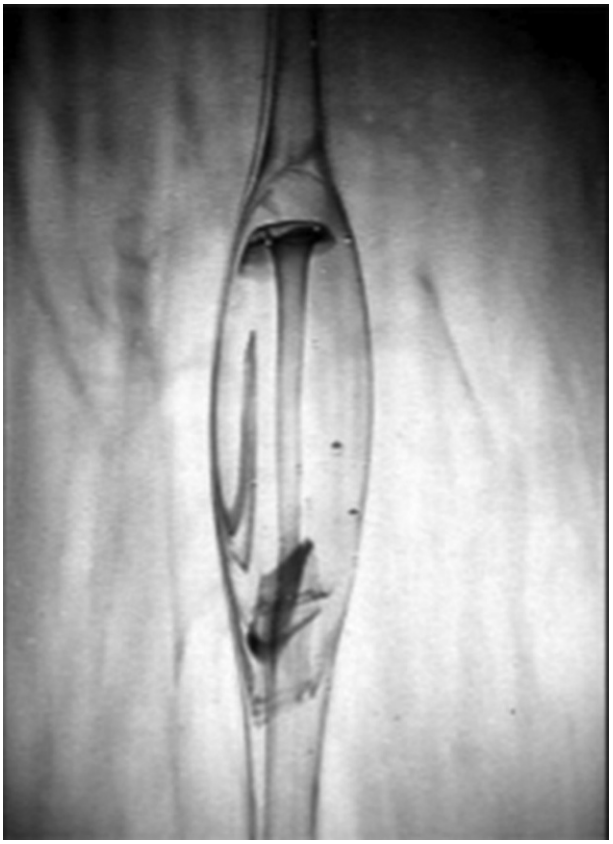


Figure 44 Thermal soliton in glycerol. Reproduced from Laudenbach N and Christensen UR (2001) An optical method for measuring temperature in laboratory models of mantle plumes. *Geophysical Journal International* 145: 528534.

latter case and then turns to the dynamics of steady plume stems and thermal starting plumes.

7.03.8.2.1 Thermals

Griffiths (1986a,b) conducted a laboratory study of thermals generated by injecting a known volume of heated polybutene oil into a large tank of the same oil at room temperature. He observed that a thermal initially rises at constant speed through a distance of the order of its diameter and then slows down while increasing in volume. Griffiths proposed that the enlargement occurs because diffusion of heat away from the thermal warms a thin boundary layer of ambient fluid around it, which is then advected back to the thermal's trailing edge and entrained into it. The total buoyancy of the thermal is therefore constant and is proportional to the Rayleigh number $Ra = \alpha g \Delta T_0 V_0 / k \nu_m$, where $V_0 = \pi a_0^3$ is the thermal's initial volume and ΔT_0 its initial temperature excess. Griffiths (1986a,b) suggested that for times greater than $\tau_G \sim Ra^{-1/2} a_0^2 / \kappa$, the thermal's volume V and velocity W have the self-similar forms:

$$V = \frac{\pi C_1^3}{6} Ra^{3/4} (\kappa t)^{3/2} \quad [23]$$

$$W = \frac{f}{2\pi C_1} Ra^{3/4} (\kappa t)^{-1/2} \quad [24]$$

where $f = O(1)$ is a function of the viscosity contrast between the thermal and the ambient fluid and $C_1 \sim 1.0$ is a constant determined experimentally. Laboratory experiments using viscous oils (Griffiths, 1986a) and corn syrup (Coulliette and Loper, 1995) are consistent with eqns [23] and [24]. In the mantle, thermals with radii >350 km would reach the lithosphere before being influenced by thermal diffusion in typically 20 My, and their ascent speeds could reach 10–20 cm year^{-1}.

7.03.8.2.2 The stem

When a starting plume rises from a steady point source of heat, the lower part of its stem quickly achieves a steady-state structure (**Figure 45(d)** and **45(e)**) that is independent of the dynamics of the ascending head above. To understand this structure, Batchelor (1954) studied analytically a model in which a plume rises from a point source of heat of strength P in an infinite fluid of constant physical properties. Simple scaling arguments suggest that the vertical velocity W, the temperature anomaly ΔT, and the stem radius R scale with the height z above the source as

$$W \sim \left(\frac{g \alpha P}{\pi \rho C_\rho \nu} \right)^{1/2} \quad [25]$$

$$\Delta T \sim \frac{P}{\kappa \rho C_P z} \quad [26]$$

$$R \sim \left(\frac{\rho C_P \nu \kappa^2}{g \alpha P} \right)^{1/4} z^{1/2} \quad [27]$$

The vertical velocity is constant, but thermal diffusion causes the temperature anomaly to decrease with height as z^{-1} and the stem radius to increase as $z^{1/2}$. Equations [25]–[27] are valid to within multiplicative functions of the Prandtl number Pr, which have been determined numerically and/or analytically (Fuji, 1963; Vasquez et al., 1996; Worster, 1986). Experiments in water ($Pr \sim 7$) verify the scalings [25]–[27] and agree with the complete functional form of Fuji (1963) if corrections are made for the finite size of the heat source and for downward heat losses from it (Shlien and Boxman, 1979; Tanny and Shlien, 1985).

The scaling [25]–[27] is no longer valid in the limit of infinite Pr, because the model problem studied by Batchelor (1954) has no solution if the fluid is infinite and inertia is totally absent (Stokes paradox; see **Chapter 7.04**). However, a solution does exist if the heat source is located on a horizontal (rigid or free) boundary and has a vertical velocity W given by eqn [25] multiplied by a function that increases logarithmically with height (Whittaker and Lister, 2006a,b). A combined experimental and numerical study in silicone oils further showed that these theoretical results were well recovered for the temperature structure but that the velocity structure was also very sensitive to the presence of the top and lateral boundaries in the experimental boxes (van Keken et al., 2013; Vatteville et al., 2009).

When the viscosity depends strongly on temperature, the thermal structure of the plume stem is observed to be well predicted by the constant-viscosity model of Batchelor (1954) for viscosity ratios up to 50. However, the particle velocity inside the stem decreases with increasing height (Davaille et al., 2011; Vatteville et al., 2009). For higher viscosity ratios, the velocity and temperature decrease can be even stronger, as seen by Laudenbach (2001) along the stem of a plume of

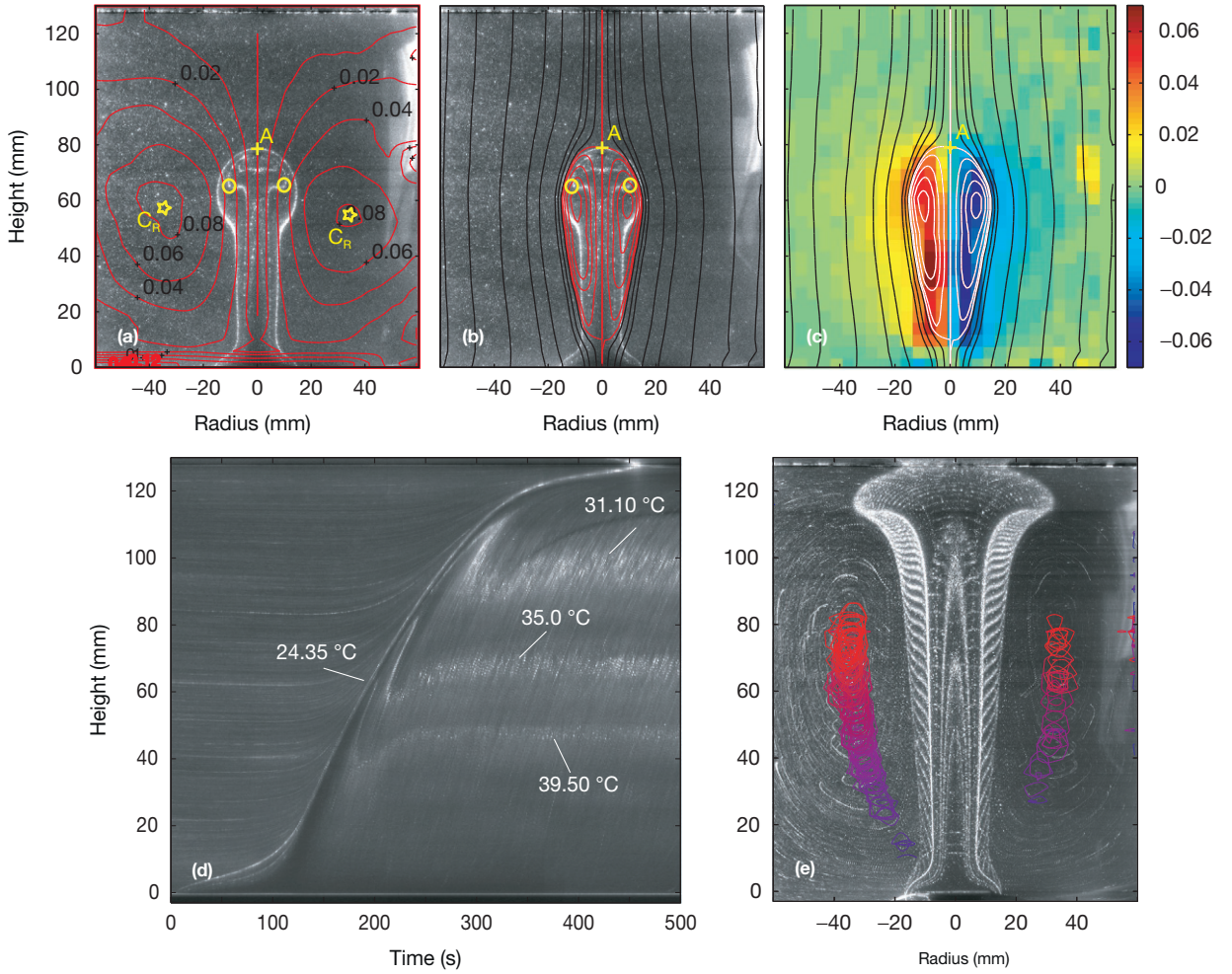


Figure 45 Thermal starting plume in sugar syrup ($P = 3$ W; $\gamma = 47$). (a) Streamlines in the laboratory reference frame on top of the plume image. The yellow cross represents the stagnation point A at the head top, the stars the rotation centers CR in the laboratory reference frame, and the yellow open circles point to the vorticity extrema. (b) Streamlines in the 'plume head' reference frame (see text) on top of the plume image and (c) on top of the vorticity field. (d) Plume uplift through time followed through the image intensity along the plume vertical axis; the bright lines are the isotherms (labeled in white). (e) Superposition of the plume images taken every 5 s between $t = 100$ and $t = 500$ s. The internal structure of the plume conduit, with its 5 isotherms, appears to be steady. The outer bright line underlines the steady growth of the plume head. The positions of the centers of rotation (corresponding to the position of the maximum values of the stream function) are represented in blue (early times) to red (late times). Adapted from Davaille A, Limare A, Touitou F, Kumagai I and Vatteville J (2011) Anatomy of a laminar starting plume at high Prandtl number. *Experiments in Fluids* 50: 285–300.

heated corn syrup, which was injected at the bottom of a tank filled with the same fluid at room temperature. These observations are in line with Olson et al.'s (1993) approximate solution of the boundary layer equations for the steady flow above a point source of heat in a fluid whose viscosity depends exponentially on temperature. The upwelling velocity W on the axis and the temperature anomaly ΔT vary with height z as

$$W = \left(\frac{g\alpha P}{4\pi\rho C_P v_0(z)}\right)^{1/2} \quad [28]$$

$$\Delta T = \Delta T_0 . \exp\left(\frac{-4\pi\kappa\Delta T_r z}{P}\right) \quad [29]$$

where $v_0(z)$ is the viscosity on the plume centerline, ΔT_0 is the temperature anomaly at the base of the plume, k is the thermal conductivity, and ΔT_r is the temperature required to change the viscosity by a factor e. The expression [28] for W is identical to the constant-viscosity expression [25] except for the variability of the centerline viscosity $v_0(z)$. The temperature anomaly decreases exponentially with height, and the corresponding increase in viscosity causes the upwelling velocity W also to decrease with height, in qualitative agreement with the experimental observations. The differences relative to the constant-viscosity case are due to the fact that temperature-dependent viscosity concentrates the upwelling in the hottest central part of the plume stem, so that the radius δ_T of the thermal anomaly is wider than the radius δ_w of the upwelling region (the 'conduit' proper) even for $Pr \gg 1$ (Figure 46). For typical mantle parameter values, plumes with large viscosity variations ($\geq$1000) would have $\delta_w \sim 25$–30 km and $\delta_T \sim 60$ km, while plumes with low viscosity variations ($\leq$100) are typically twice as broad.

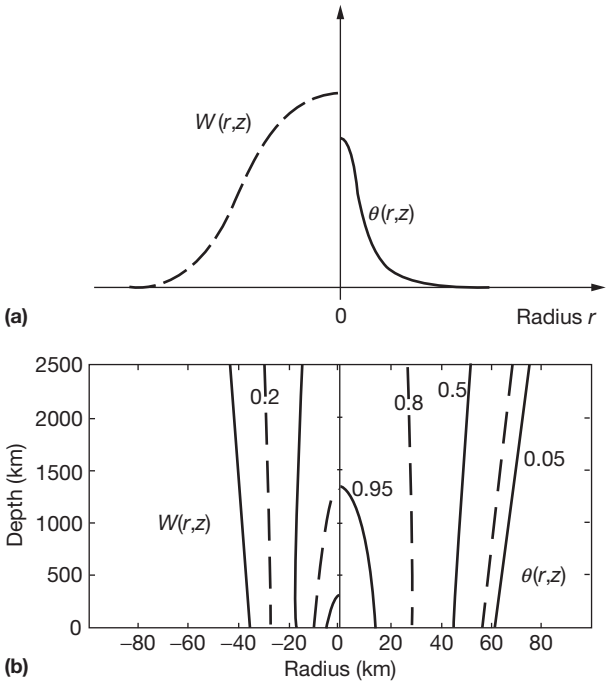


Figure 46 Thermal plume stem at high-Prandtl number. (a) Sketch of the radial temperature (solid line) and vertical velocity (dashed line) profiles for constant viscosity: the plume thermal anomaly is embedded into the velocity boundary layer. (b) Model of Olson et al. (1993) for strongly temperature-dependent viscosity in the mantle: the high-velocity conduit (left) is embedded in a 'thermal halo' (right). Temperature and velocity are normalized by their maximum value on the axis.

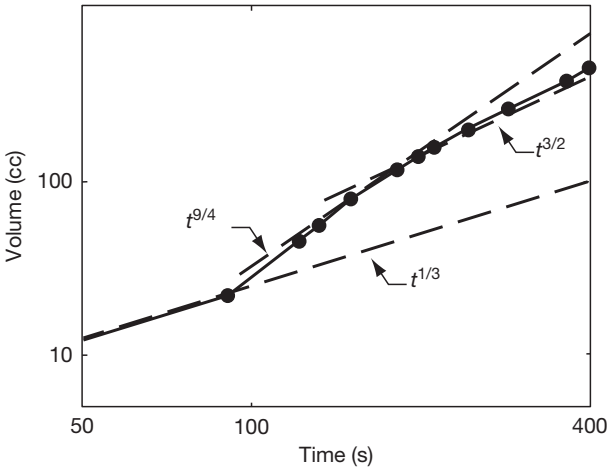


Figure 47 Volume of a starting plume as a function of time. The circles represent the experimental data points from Griffiths and Campbell (1990). In dashed lines, the $t^{1/3}$ law for a growing Stokes sphere (Whitehead and Luther, 1975), the $t^{9/4}$ growth model of Griffiths and Campbell (1990), and the $t^{3/2}$ law due to the conductive growth of the plume cap.

7.03.8.2.3 Thermal starting plumes

A thermal starting plume is essentially an ascending thermal connected by a stem to a source of buoyancy (**Figure 45**). Thus, the plume head can grow both by the entrainment of ambient fluid and by the addition of plume fluid through the stem. Unfortunately, different studies seem to give conflicting results regarding the evolution of the size and speed of the plume head and the initial location of the material sampled by the plume.

Griffiths and Campbell (1990) investigated thermal starting plumes by injecting, at a constant volumetric rate Q, hot glucose syrup with a temperature excess ΔT_0 into a cooler reservoir of the same fluid. They proposed that the volume V of the plume head evolves as

$$\frac{dV}{dt} = Q + C_2 \left(\frac{\kappa g \alpha \Delta T_0}{\nu_m}\right)^{1/2} V^{1/3} \qquad [30]$$

where ν_m is the viscosity of the cold ambient fluid and C_2 is an empirical entrainment constant. For large times, the second (entrainment) term in eqn [30] is dominant, implying

$$V = \left(\frac{4C_2}{9}\right)^{3/2} \beta \kappa^{3/4} t^{9/4} \qquad [31]$$

where $\beta = (\alpha g \Delta T_0 Q/\nu_m)^{3/4}$. The corresponding velocity of the head, calculated from Stokes law for a relatively inviscid drop, is

$$W = \frac{1}{8}\left(\frac{6}{\pi}\right)^{2/3} C_2^{-1/2} \beta (\kappa t)^{1/4} \qquad [32]$$

On the basis of experiments with thermals, Griffiths and Campbell (1990) stated that $1 \leq C_2 \leq 4$. No velocity data were provided. Equation [31] describes well the head volume experimental data for intermediate times, although at larger times, the cap growth rate becomes lower than predicted and is better described by the law $V \sim t^{3/2}$ eqn [23] corresponding to the much slower growth by thermal diffusion of an isolated thermal (**Figure 47**). However, **Figure 7(c)** with transparent ambient fluid embedded into the prominent scroll of injected colored fluid, as well as eqn [31], conveys the image of a plume head rapidly containing a large proportion of entrained material.

Experiments on plumes produced by localized heating without fluid injection show a behavior quantitatively different. First, the velocity of the head increases rapidly for a short time and then either decreases to reach or attains directly a nearly constant value W_{term} (Coulliette and Loper, 1995; Davaille et al., 2011; Kaminski and Jaupart, 2003; Lohmann et al., 2009; Moses et al., 1993; Vatteville et al., 2009). To explain this, Coulliette and Loper (1995) proposed a modified version of Griffiths and Campbell (1990)'s model that accounted for incomplete entrainment of the hot TBL surrounding the head. However, the resulting theory involves empirical constants whose values vary by factors of 2–5 among experiments. A possible cause of the difficulty may be the theory's neglect of the flux required to build the lengthening stem, which is important for compositional plumes (Olson and Singer, 1985; Whitehead and Luther, 1975). A second observation differs from the injection experiments: the plume head growth is much slower than predicted by eqn [31] (**Figure 45(e)**). Recent techniques allow to follow the ascent and growth of a starting plume through both its velocity and temperature fields. Key measurements are then the velocity of the stagnation point located on the top of the plume head V_A and the velocity of the particles at this point $V_z(A)$. The former will give the same information as following dye (which underlines the maximum deformation) or following shadowgraph (which underlines the maximum temperature gradient), while

$V_Z(A)$ gives the same value as the uplift of the maximum vorticity centers. The latter constitutes the 'true' plume head velocity. The difference $V_A - V_Z$ (A) represents plume head growth by thermal diffusion (**Figure 45(a)–45(c)**; Davaille et al., 2011). Moreover, the plume samples only the material coming from the TBL above the heat source (**Figure 45(b)**; Androvandi et al., 2011; Davaille et al., 2011; Lohmann et al., 2009), in agreement with early numerical simulations (Farnetani and Richards, 1995).

7.03.8.3 Plumes in Non-Newtonian Fluids

We have already mentioned in **Section 7.03.6.3** that viscoplastic rheology was often used in numerical simulations to produce weak zones. Recent experiments have investigated the generation of thermal plumes in Carbopol, a viscoelastoplastic fluid (see **Section 7.03.2.2**; **Figure 4**). In the latter, the morphology of plumes is very different from the mushroom-shape observed in the previous section: a plug flow develops inside the plume thermal anomaly, producing a rising finger-shape with strong shear zones confined along its edges (**Figure 48**). And two conditions must now be fulfilled for an instability to develop and rise: (1) The yield number ψ_0, comparing the buoyancy-induced stress to the yield stress, should be greater than a critical value $\psi_0 \sim 6$ (Davaille et al., 2013); and (2) the Bingham number Bi, comparing the yield stress to the viscous stresses, needs to be locally smaller than 1. Moreover, the instability halts its ascent as soon as $\psi_0 < \psi_c$ or $Bi > 1$ (Massmeyer et al., 2013).

The existence of >100 km size mushroom-shaped mantle plumes then implies that $\sigma_0 \ll 100$ MPa in the bulk mantle. However, the experimental finger-shaped plumes show strong similarities with diapiric intrusions in the Earth's lithosphere, such as documented on the field in Oman (Jousselin and Nicolas, 2000) or inferred from uplift in the Altiplano–Puna region (Fialko and Pearse, 2012). But a 10 km diameter purely thermal anomaly could only be emplaced in a partially molten lithosphere (with $\sigma_0 \sim 100$ kPa). The yield stress of the surrounding matrix might be much higher (up to $\sigma_0 \sim 50$ MPa), if the instability also exhibits a chemical density anomaly. On the other hand, this critical yield stress constitutes a lower value of σ_0 for an intrusion to remain trapped in the lithosphere (Massmeyer et al., 2013).

7.03.8.4 Interaction of Plumes with a Large-Scale Flow

Because deep-seated plumes must traverse the convecting mantle on their way to the surface, it is important to investigate how they interact with an ambient large-scale flow. This interaction was first studied by Skilbeck and Whitehead (1978), who injected compositionally buoyant fluid into a shear flow generated by rotating the top surface of a tank containing another fluid. As the rising plume head is swept horizontally away from the source, the trailing conduit is tilted until it becomes gravitationally unstable at a tilt angle 30° from the vertical. The instability forms new, smaller cavity plumes whose spacing is independent of the volume flux Q (Whitehead, 1982). Olson and Singer (1985) performed similar experiments by towing the source at constant speed U at the bottom of a tank of motionless fluid (**Figure 20(d)**). The characteristic spacing λ and volume V of the new cavity plumes generated by the instability were found to scale as $\lambda \sim 12(\nu_m U/g^*)^{1/2}$ and $V \sim 12Q(\nu_m/g^* U)^{1/2}$, respectively, although these relationships break down for small (<0.3) and large (>10) values of the dimensionless parameter $(g^*Q/\nu_m U^2)^{1/2}$. Skilbeck and Whitehead (1978) suggested that this

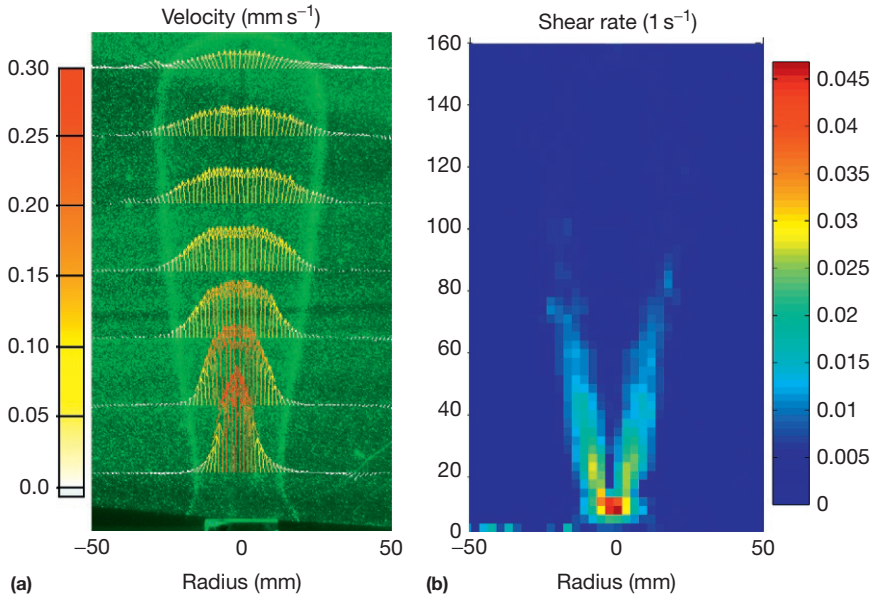


Figure 48 Thermal plume in a yield stress and shear-thinning Carbopol gel. (a) Temperature (visualized by isotherms) and velocity structure. (b) Shear rate calculated from the velocity fields. Note how deformation is confined on the edges of the thermal finger. Reproduced from Davaille A, Gueslin B, Massmeyer A, Di Giuseppe E (2013) Thermal instabilities in yield stress fluids: existence and morphology. *Journal of Non-Newtonian Fluid Mechanics* 193: 144–153.

instability could produce a chain of discrete volcanic islands. However, their typical spacing would be ~30–100 km for a plate speed of ~10 cm year^{-1} and a mantle viscosity of 10^{20}–10^{21} Pa s, and those values are smaller than what is observed. The spacing of volcanic island chains is probably rather controlled by lithospheric flexure and magma transport (e.g., Hieronymus and Bercovici, 1999; Ten Brink, 1991).

Richards and Griffiths (1988) studied experimentally the interaction between a compositionally buoyant plume conduit and a large-scale shear flow, for low conduit tilt angles at which the instability documented by Skilbeck and Whitehead (1978) does not occur. They showed that the conduit's shape can be calculated by describing its motion as a vector sum of the shear velocity and a modified vertical Stokes velocity v_S for individual conduit elements. For a simple linear shear flow, the resulting steady-state shape is a parabola. In response to a sudden change in the shear flow, the lateral position $x(z)$ of such a conduit adjusts to a new steady-state position in a time z/v_S, where z is the height. Griffiths and Richards (1989) applied this theory to Hawaii, concluding that the sharpness of the bend in the Hawaiian–Emperor chain at 43 Ma implies that the deflection of the Hawaiian plume before the change in plate motion was <200 km. Kerr and collaborators (Kerr and Mériaux, 2004; Kerr et al., 2008) carried out a comprehensive laboratory study of thermal plumes ascending in a shear flow in a fluid with temperature-dependent viscosity. Plumes were observed to rise initially with a constant velocity that is independent of the centerline viscosity v_P. Subsequently, the ambient shear tilts the plume conduits, and in some cases, significant cross-stream circulation and entrainment are seen. However, the tilted conduits never develop Rayleigh–Taylor-type gravitational instabilities of the kind observed on tilted compositional conduits (**Figure 20(c)** and **20(d)**), because thermal diffusion is fast enough to suppress their growth (Kerr et al., 2008; Richards and Griffiths, 1988). Moreover, some of the buoyancy flux in mantle plumes can be lost during ascent to form downstream thermal wakes in any larger-scale mantle flow (Kerr et al., 2008).

7.03.8.5 Plume–Lithosphere Interactions

We consider here the close plume–lithosphere interactions occurring when a mantle plume impinges under the lithosphere. Since the interaction of plumes with the lithosphere is responsible for most of the geochemical and geophysical signatures observed at hot spots, the dynamics of this process has been the focus of numerous experimental studies. The more general case of the interaction of plumes with density and/or viscosity boundaries within the mantle will be treated together with two-layer convection in **Section 7.03.9**.

7.03.8.5.1 Stagnant lithosphere

To understand the formation of bathymetric swells by plumes, Olson and Nam (1986) performed laboratory experiments on the deformation of an initially spherical drop of buoyant low-viscosity fluid rising toward a high-viscosity boundary layer at the cooled free surface of a fluid with strongly temperature-dependent viscosity. They measured the amplitude of the surface topography as a function of time using a proximity probe (**Section 7.03.3.5**; **Figure 19(a)**) and found that the typical evolution comprises a phase of rapid uplift followed by slower subsidence that corresponds to stagnation and lateral spreading of the drop below the cold boundary layer. The maximum swell height increased with the drop volume but decreased with increasing lithosphere thickness. Griffths et al. (1989) confirmed these results using holographic interferometry (**Figure 19(c)**) and extended the study to larger viscosity and buoyancy contrasts between the lithosphere and the underlying mantle. They found the maximum swell height to be independent of the viscosity contrast, but a decreasing function of the density contrast.

Using experiments performed under isothermal conditions, Griffiths and Campbell (1991) determined scaling laws for the radius of a drop spreading beneath a solid plate and the thickness of the squeezed film above the drop as functions of time. They observed that the gravitationally unstable interface between the drop and the squeeze film eventually broke up via a RTI with a complex lateral planform dominated by short-wavelength (~drop thickness) upwellings and downwellings.

Olson et al. (1988) studied the penetration of mantle plumes through colder and more viscous lithosphere, while Jurine et al. (2005) investigated how thermal plumes deform and penetrate a compositionally distinct lithosphere. The amplitude and shape of the penetration were found to depend mostly on the ratio B between the compositional and thermal density contrasts ('buoyancy number'; see **Section 7.03.9**), while the timescale of the interface displacement is governed primarily by the viscosity contrast. When applied to the continental lithosphere, the experimentally derived scaling laws show that plume penetration through Archean lithosphere requires thermal anomalies larger than 300 °K, corresponding to buoyancy fluxes in excess of those determined today for the Hawaiian plume. But because the experimental plumes are observed to interact differently with lithospheres of different ages and/or compositions, no single 'typical' plume signature can be identified.

7.03.8.5.2 Moving or rifting lithosphere

While the interaction of a plume with a stationary lithosphere is relatively easy to study in the laboratory, analogue modeling of a moving lithosphere poses serious challenges, including the difficulty of controlling the temperature on the moving boundary and the unwanted influence of the return flow the boundary motion induces (see **Section 7.03.7.**) Here, two cases must be distinguished: interaction of the plume with an intact lithosphere moving at constant speed relative to it (plume–plate interaction or PPI) and interaction of the plume with a rifting lithosphere representing a mid-ocean ridge (plume–ridge interaction or PRI). To date, there exist no experimental studies of PPI, which has been investigated solely with analytic and numerical methods (e.g., Olson, 1990; Ribe and Christensen, 1994, 1999). By contrast, PRI has been successfully modeled in the laboratory.

The first case to be studied was that of a 'ridge-centered' plume rising directly beneath a spreading ridge, Iceland being the main natural example. Feighner and Richards (1995) investigated this case by injecting compositionally buoyant fluid from a pipe beneath the boundary between two diverging mylar sheets at the top of a tank of corn syrup. The buoyant

fluid is observed to spread beneath the mylar sheets to form a thin layer with a steady 'bowtie' shape in map view. And the width of the layer along the ridge scales as $(Q/U)^{1/2}$, where Q is the volumetric injection rate and U is the half-spreading rate, in agreement with the predictions of numerical models based on lubrication theory (Ribe et al., 1995) and the tracking of chemically buoyant tracer particles (Feighner et al., 1995). This work was extended by Kincaid et al. (1995) to the interaction of a thermal plume with a mylar-sheet spreading ridge located at some distance from it. The mantle was modeled using a concentrated sucrose solution, producing a highly viscous upper boundary layer (the 'lithosphere') that cooled and thickened as it moved away from the ridge. Kincaid et al. (1995) observed that the plume material can be guided toward the ridge along the sloping base of the lithosphere, confirming the suggestion of Schilling (1991) that an off-axis plume may communicate thermally and chemically with a spreading ridge.

7.03.9 Inhomogeneous Fluids: Mixing and Thermochemical Convection

These studies have arisen from two fundamental questions in mantle dynamics:

(a) How can the dispersion observed in geochemical data be interpreted in dynamic terms? Since geochemical isotopic signatures modify neither the density nor the physical properties (rheology, thermal expansion, and thermal diffusion) of mantle materials, they constitute passive tracers of mantle flow. The question therefore relates to the more general problem of convective mixing of passive tracers.

(b) What is the convective pattern in the mantle? Can mantle convection be stratified? There are numerous sources of compositional heterogeneities in the mantle, such as slab remnants (e.g., Christensen and Hofmann, 1994; Olson and Kincaid, 1991), delaminated continental material, relics of a primitive mantle (Gurnis and Davies, 1986; Kellogg et al., 1999; Tackley, 1998) enriched, for example, in iron (Javoy, 1999), and chemical reactions with or infiltration from the core (e.g., Hansen and Yuen, 1988). So, how does a density heterogeneity interact with thermal convection?

Both questions involve time-dependent 3-D phenomena occurring on a wide range of spatial scales, abrupt gradients of physical properties (e.g., density and viscosity), and intricate physical phenomena (stirring and entrainment) for which even the equations remain to be found. Laboratory experiments have therefore proved to be a good tool to tackle these problems. A more complete review of mixing in the mantle is given by Tackley in **Chapter 7.12**.

7.03.9.1 Mixing of Passive Tracers

An impurity is said to be a passive tracer when it does not modify the flow field transporting it. This condition requires the impurity either to be neutrally buoyant or to have a characteristic length scale very small compared to that of the flow. Mixing of an impurity involves two phenomena: (a) first stirring, that is, stretching and folding of the heterogeneity until its length scale becomes small enough for chemical diffusion to act, and (b) diffusive erasure of the heterogeneity (e.g., Nagata, 1975; Ottino, 1989; Reynolds, 1894). Given the small values of chemical diffusion coefficients in the mantle (typically 10–15 $cm^2\ s^{-1}$; Hofmann and Hart, 1978), the latter process will act only over a few cm on geologic timescales. Stirring is therefore the process of greater interest for geodynamics (e.g., Allegre, 1987; see also Ricard, **Chapter 7.02**).

Our current understanding of mixing has significantly improved in the last 20 years, due to the establishment of a clear connection between chaos and stirring (e.g., Aref, 1984). Essential to this breakthrough were careful laboratory experiments in prototypical flows, first in 2-D, time-periodic flows (e.g., Ottino et al., 1988) and then in 3-D flows (e.g., Fountain et al., 1998). In particular, studies in simple periodic 2-D flows have helped visualize chaos (**Figure 49**): unmixed regions translate, stretch, and contract periodically and represent the primary obstacle to efficient mixing. Particle trajectories in chaotic regions separate exponentially in time, and material filaments are continuously stretched and folded by means of horseshoes. The details seen in **Figure 49** have long been impossible to recover in computer simulations since numerical diffusion was greater than the molecular diffusion in the fluid experiments. However, recent developments in numerical methods (see Tackley, this volume) and improvements in computer power are quickly filling the gap. Moreover, a special effort has been made in mixing studies to couple to the extent possible simple lab experiments and numerical simulations of the same flow, since the latter give much more ready access to the velocity field and other quantities such as the time-dependent stretching field. Recent theoretical work has provided a solid connection between maxima of the stretching field and invariant manifolds of the flow (Haller, 2001; Haller and Yuan, 2000). In a flow that can be described analytically,

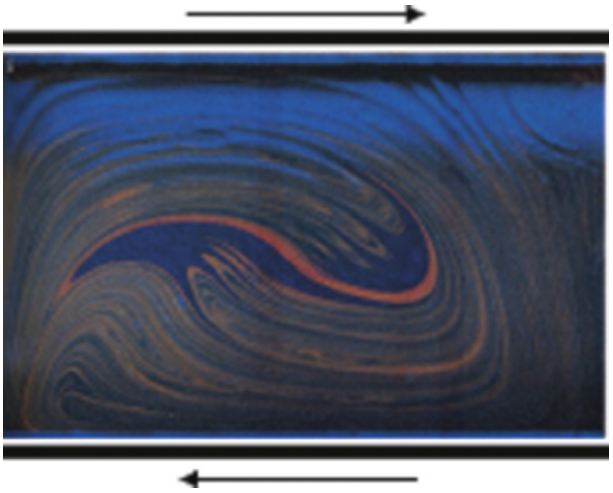


Figure 49 Mixing produced by discontinuous cavity flow. The top wall moves from left to right for half a period and then stops, and then, the bottom wall moves from right to left and stops, and so on. Visualization is provided by a fluorescent tracer dissolved in glycerine excited by a long-wave ultraviolet light of 365 nm. Folding and an island of unmixed material are clearly seen. Reproduced from Ottino JM, Leong CW, Rising H, and Swanson PD (1988) Morphological structures produced by mixing in chaotic flows. *Nature* 333: 419–426.

the comparison between computations of stretching and dye visualization shows that the dye spreads over regions that have experienced large stretching. The use of high-resolution PIV technique in laboratory experiments now allows to calculation of the in situ stretching field (e.g., Voth et al., 2002). This gives a powerful new insight into the geometric structure that underlies mixing, even for nonanalytic flow field.

The more specific study of mixing by thermal convection in the laboratory has been confined to 2-D and weakly 3-D patterns. Solomon and Gollub (1988a,b) studied the passive transport of methylene blue and latex spheres by steady 2-D Rayleigh–Bénard convection. The convective flow was monitored by LDV and the transport by optical absorption techniques. The transport in this system is determined by the interplay between advection of tracer particles along the streamlines within convection rolls and diffusion of these particles between adjacent rolls. They found that the transport on long space and timescales could be modeled as a diffusive process with an effective diffusion coefficient $D^* \sim DPe^{1/2} = D(W_d/\pi D)^{1/2}$ where D is the molecular diffusion, W_d the characteristic velocity of the flow, and Pe the Péclet number. When the 2-D convection is time-periodic, the local effective diffusion coefficient becomes independent of the molecular diffusion but depends linearly on the local amplitude of the roll oscillation. The basic mechanism of transport is chaotic advection in the vicinity of oscillating roll boundaries (Solomon and Gollub (1988a,b)). Solomon et al. (1996) verified that D^* is related to the areas of the lobes that are carrying tracers between two cells (**Figure 50**). For weakly time-dependent and 3-D flows, Solomon and Mezic (2003) observed completely uniform mixing when the pattern oscillation period was close to typical circulation times, as predicted by 'singularity-diffusion' theory.

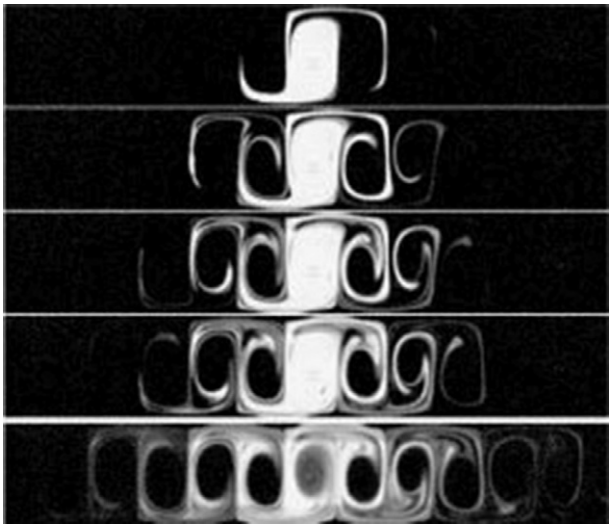


Figure 50 Transport of uranine dye along an array of convection cells oscillating with a amplitude of 0.12. Time (from the top): 1, 2, 3, 4, 10 oscillations. Reproduced from Solomon TH, Tomas S, and Warner JL (1996) The role of lobes in chaotic mixing of miscible and immiscible impurities. *Physical Review Letters* 77: 2682–2685.

7.03.9.2 Convection in an Initially Stratified Fluid

The most commonly studied situation comprises two superposed fluid layers of different composition, density, and viscosity (**Figure 51**) across which a temperature difference ΔT is applied. Since convection occurs in the Earth's mantle on large scales ($\gg 1$ km), and surface tension acts on mm scales (i.e., laboratory scales), we shall confine our review to the case of two miscible fluids. When the viscosity depends on temperature, ν_m and ν_d are taken to be the values at the mean temperature of each layer. In cases where the thermal expansion coefficient also depends on temperature (Le Bars and Davaille, 2004a,b), it is taken to be the value at the temperature of the layer interface.

Accordingly, the system dynamics are characterized by five dimensionless numbers:

- The Rayleigh number, Ra (H, ΔT), defined using eqn [1] with $\nu = \max(\nu_d, \nu_m)$
- The Prandtl number $Pr = \nu/\kappa$
- The viscosity ratio γ
- The depth ratio $a = h_d/H$
- The buoyancy ratio, $B = \Delta\rho_\chi/\rho\alpha\Delta T$, ratio of the stabilizing chemical density anomaly $\Delta\rho_\chi$ to the destabilizing thermal density anomaly

7.03.9.2.1 Regime diagrams

Early work used linear stability analysis to investigate the stability of various two-layer systems as a function of Rayleigh number and density contrast (Busse, 1981; Richter and Johnson, 1974; Sotin and Parmentier, 1989). The presence of a deformable interface offers the possibility of Hopf bifurcations: since it introduces an additional degree of freedom, overstability can occur in the form of oscillatory interfacial instability (Jaupart et al., 2007; Le Bars and Davaille, 2002; Rasenat et al., 1989; Renardy and Joseph, 1985a,b; Renardy and Renardy, 1985; Richter and Johnson, 1974). This overstable mode was observed experimentally for the first time by Le Bars and Davaille (2002) (**Figure 52**). At first, numerical studies of two-layer convection at finite amplitude have typically focused on the mechanism of coupling (thermal vs. mechanical) between the layers, often treating the interface as impermeable (e.g., Boss and Sacks, 1986; Cserepes and Rabinowicz, 1985; Cserepes et al., 1988; Christensen and Yuen, 1984; Kenyon and Turcotte, 1983; Richter and McKenzie, 1981). For a long time, the highly nonlinear convective regimes associated with high Rayleigh numbers or entrainment at the interface were accessible only to laboratory experiments.

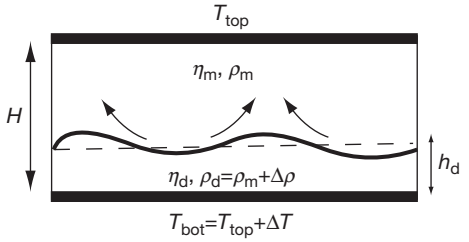


Figure 51 Sketch of two-layer convection. The denser layer is initially at the bottom.

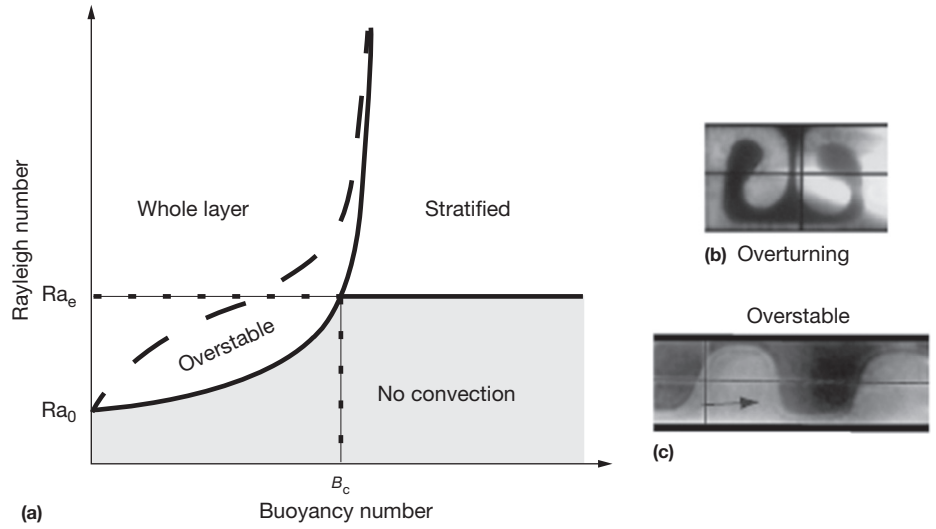


Figure 52 (a) Marginal stability curves separating the different convective regimes. The curves here were calculated for the case $\gamma = 6.7$ and $a = 0.5$, where $Ra_0 = 5430$, $Ra_c = 38\,227$, and $B_c = 0.302$. For $B < B_c$, the whole-layer regime develops under the form of (b) overturn for γ around 1 and/or low B (here, $B = 0.048$; $Ra = 6.7 \times 10^3$; $a = 0.44$; $\gamma = 1.1$) or (c) traveling waves at low Rayleigh number (here, $B = 0.20$; $Ra = 1.8 \times 10^4$; $a = 0.5$; $\gamma = 6.7$). Reproduced from Le Bars and Davaille A (2002) Stability of thermal convection in two superimposed miscible viscous fluids. *Journal of Fluid Mechanics* 471: 339–363.

Using aqueous glycerol solutions, Richter and McKenzie (1981) showed that two-layer convection is stable when the ratio of chemical to thermal buoyancy B exceeds unity, but they did not address the issue of partial mixing. Olson (1984) used aqueous sugar solutions to study two-layer convection in the penetrative regime $B \sim 1$ when the viscosity ratio between the two layers is lower than 10. His results established that entrainment at low Reynolds number can indeed occur and that viscous stresses can substitute for inertial instabilities as a mixing mechanism. Nataf et al. (1988) demonstrated that the mechanism of coupling between immiscible fluid layers depends on conditions at the interface (e.g., surface tension and impurities) and that the oscillatory interfacial mode is stabilized by surface tension, as predicted by theory.

Numerous studies now exist on the subject, and the domain of parameters investigated is therefore large, with $10^2 < Ra < 10^9$, $10^{-2} < \gamma < 6 \times 10^4$, $0.05 < a < 0.95$, and $0.04 < B < 10$. The experiments have been run with constant viscosity (e.g., aqueous Natrosol solutions) or temperature-dependent viscosity (e.g., sugar or corn syrups) fluids and with constant or temperature-dependent thermal expansion. The different regimes observed not only in laboratory experiments (Cottrell et al., 2004; Davaille, 1999a,b; Davaille et al., 2002; Jaupart et al., 2007; Jellinek and Manga, 2002, 2004; Le Bars and Davaille, 2002, 2004a; Olson, 1984; Olson and Kincaid, 1991; Richter and McKenzie, 1981) but also in numerical simulations (Deschamps et al., 2012; Hansen and Yuen, 2000; Kellogg et al., 1999; McNamara and Zhong, 2004a; Montague and Kellogg, 2000; Samuel and Bercovici, 2006; Samuel and Farnetani, 2002; Schmeling, 1988; Tackley, 1998, 2002; Tan et al., 2008) are shown in **Figure 53** and **Table 1** and described in more detail later.

For $B < 0.03$, chemical density heterogeneities are negligible and the system behaves as a single homogeneous fluid (**Figure 53(a)**). For larger values of B and $Ra > 10^5$, two scales of convection coexist: compositionally homogeneous thermal plumes generated at the outer boundaries and large-scale thermochemical instabilities that involve both fluids.

7.03.9.2.2 Whole-layer regime and unstable doming

When B is less than a critical value $B_c \sim 0.4$ (the exact value depending on the viscosity and depth ratios; Jaupart et al., 2007; Le Bars and Davaille, 2002), the stable compositional stratification can be overcome by thermal buoyancy. The interface between the layers then becomes unstable, and convection occurs over the whole depth. An important characteristic of the interfacial instability is the spouting direction (Whitehead and Luther, 1975), defined as the direction (up or down) in which finite-amplitude perturbations grow superexponentially to form dome-like structures (**Figure 53(e)** and **53(f)**). The theory of the RTI shows that spouting occurs in the direction of the layer with the lower resistance (Ribe, 1998). Thus, the lower layer will spout into the upper one only if

$$h_d < \frac{H}{1 + \gamma^{-1/3}} \qquad [33]$$

Otherwise, spouting is downward (Le Bars and Davaille, 2004a). Condition [33] implies, for example, that a less viscous layer will spout into a more viscous overlying mantle only if the former is much thinner. As for the case of purely compositional plumes (**Figures 7, 20(a)**, and **20(b)**), the morphology of thermochemical instabilities depends on the viscosity ratio γ. Suppose for definiteness that eqn [33] is satisfied, so that spouting is upward. Then, when the lower layer is more viscous ($\gamma > 1$), large cylindrical diapirs form (**Figure 53(f)**), and secondary plumes can develop above them in the upper layer (Davaille, 1999b). When $\gamma < 1$, by contrast, purely thermal instabilities first develop within the lower layer and then merge to form large cavity plume heads

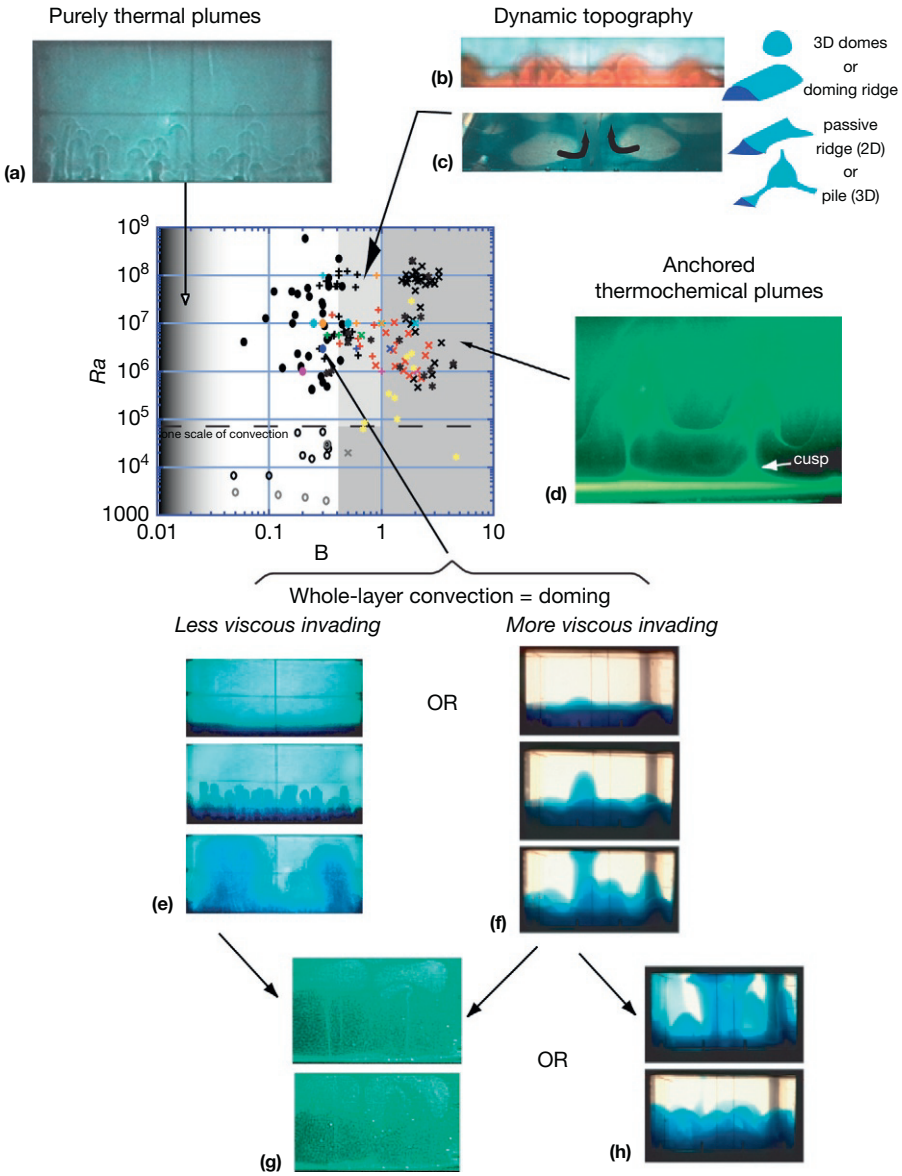


Figure 53 Regimes of thermochemical convection as a function of Rayleigh number and buoyancy number. Black circles: whole-layer convection (open circles for experiments close to marginal stability), +: stratified convection and anchored plumes with a deformed interface (light shaded area), x: stratified convection with a flat interface, *: stratified convection and anchored plumes with a nonconvecting thinner layer. Laboratory experiments are shown in red (Richter and McKenzie, 1981), dark blue (Olson and Kincaid, 1991), yellow (Jellinek and Manga, 2004), and black (Davaille, 1999a,b; Davaille et al., 2002; Le Bars and Davaille, 2002, 2004a.) Numerical calculations are shown in gray (Schmeling, 1988), light blue (Tackley, 1998, 2002), brown (Kellogg et al., 1999; Montague and Kellogg, 2000), purple (Hansen and Yuen, 2000), green (Samuel and Farnetani, 2002), and pink (McNamara and Zhong, 2004a,b).

(Figure 53(e)). In both cases, if the lower layer is thin, the plumes empty it before they reach the upper boundary and become disconnected from the lower boundary as in the purely thermal case (Figure 29). If the layer is thicker and/or $0.2 < B < B_c$, the plumes reach the upper boundary before disconnecting from the lower (Figure 53(e)). They then begin to cool and lose their thermal buoyancy and eventually collapse back to the bottom, whereupon the cycle begins again. When $\gamma > 5$ or $\gamma < 0.2$, each layer retains its identity over several pulsations. Since thermal buoyancy must overcome the stable compositional stratification before driving convection, the temperature anomaly θ carried by thermochemical instabilities is the slave of the compositional field (Le Bars and Davaille, 2004a) and is

$$\theta = (0.98 \pm 0.12)\frac{\Delta\rho_\chi}{\alpha\rho} \qquad [34]$$

Moreover, the diameter, wavelength, velocity, and cyclicity of the domes are mainly controlled by the more viscous upper layer since it retards motion over the whole depth (e.g., Herrick

Table 1 Convective regime and upwellings morphology as a function of B, viscosity ratio γ, layer depth ratio a, and internal Rayleigh number of the denser bottom layer Ra_d

B	Regime	$\gamma = \eta_d/\eta_m$	Ra_d	$a = h_d/H$	Upwellings morphology	Figure
<0.03	1-layer				Thermal plumes	
$0.03 < B < B_c \sim 0.4$	Whole layer				Active domes and passive ridges	
			$<Ra_c \sim 1000$		Passive ridges = return flow to downwellings	53c
		<1	$>Ra_c$	$a < a_c$	Active hot upwellings	
				$a_c = 1/(1+\gamma^{-1/3})$	• Cavity plumes (or 'mega-plumes') through collection of small thermal instabilities	53e
					• detach from hot bottom boundary (HBB)	53 g
			$>Ra_c$	$a > a_c$	Passive ridges	53c
					• return flow to cold more viscous downwellings	
		>1	$>Ra_c$		• Active hot diapirs detach from HBB if $a < 0.3$ and $B < 0.2$ continuous fingers from HBB otherwise	53f 53 g 53 h
					• Secondary plumes on top of domes	
		$1/5 < \gamma < 5$			Overturning = immediate stirring after first instabilities	
		$\gamma < 1/5$ or $\gamma > 5$			Pulsations = two layers retain their identity for several doming cycles	53f + 53 h
$> B_c$	2-layers				• Stratified convection above and below interface	53b–d
					• Anchored hot thermochemical plumes arise from TBL at the interface.	53d
$B > 1$					Nearly flat interface	
$B_c < B < 1$					Dynamic topography, does not reach the upper boundary	53b–c
		< 1			Passive ridges (2D) or piles (3D) formed in response to cold viscous downwellings	53c
		>1	$<Ra_c$		Passive ridges (2D) or piles (3D)	53c
			$Ra_c < Ra_d < 10^4$		Upwelling ridges	53b
			$> 10^4$		Upwelling domes, or superplumes	53b

For depth- or temperature-dependent properties, the viscosities are taken at the averaged temperature of each layer, and α is taken at the interface. We focus on high global Ra ($>10^5$). Note that if $Ra_d < Ra_c$, the denser layer cannot convect on his own.

and Parmentier, 1994; Olson and Singer, 1985; Whitehead and Luther, 1975).

7.03.9.2.3 Stratified regime: anchored hot spots, bumps, ridges, and piles

For $B > B_c$, thermal buoyancy cannot overcome the stable compositional stratification. Convection remains stratified, and a TBL forms at the interface from which long-lived thermochemical plumes are generated (**Figure 53(d)**). These plumes do not have a well-defined head, and the thermal anomaly they carry is weak, proportional to the temperature difference across the unstable part of the TBL above the interface (Christensen, 1984; Farnetani, 1997; Tackley, 1998). They entrain a thin filament (at most 5% of the total plume volume) of the denser bottom layer by viscous coupling and locally deform the interface into cusps (**Figure 53(d)**, Davaille, 1999b; Jellinek and Manga, 2002, 2004). The interfacial topography and temperature anomaly serves in turn to anchor the plumes (Davaille, 1999a,b; Davaille et al., 2002; Jellinek and Manga, 2002, 2004; Matsumoto et al., 2006; Namiki and Kurita, 1999), which persist until the chemical stratification disappears through entrainment.

When $B_c < B < 1$, thermal buoyancy is sufficient to maintain local thermochemical 'bumps' on the interface (**Figure 53(b)** and **53(c)**), whose maximum height increases with increasing Ra and decreasing B (**Figure 54**; Cottrell et al., 2004; Jaupart et al., 2007; Kumagai et al., 2008; Le Bars and Davaille, 2004a; McNamara and Zhong, 2004a,b). If the lower layer is thin

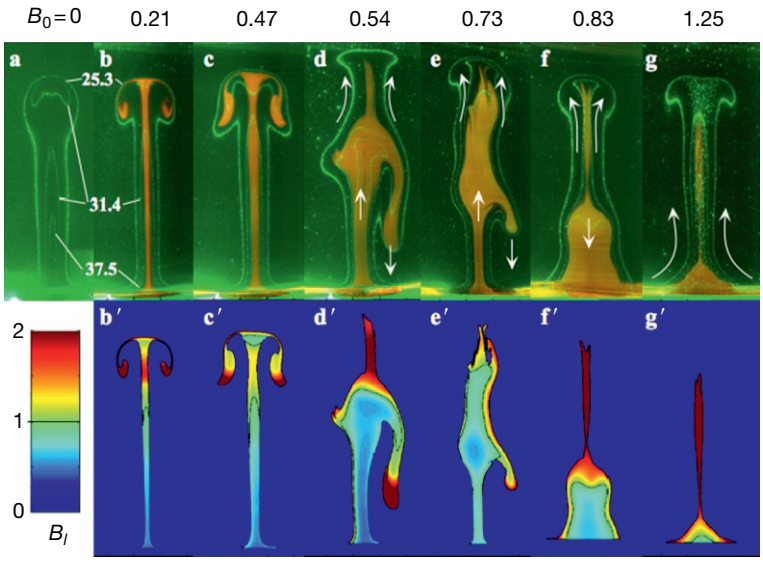


Figure 54 Different dome morphologies as B increases. A thermochemical plume was generated on a flat heater with a constant power. Initially, the orange denser material was in a thin layer at the bottom of the tank. The green lines are the isotherms. (top) Thermal and compositional fields. The white arrows indicate the fluid motions. (bottom) Local buoyancy ratio. Reproduced from Kumagai I, Davaille A, Kurita K, and Stutzmann E (2008) Mantle plumes: Thin, fat, successful, or failing? Constraints to explain hot spot volcanism through time and space. *Geophysical Research Letters* 35: L16301.

enough, it can break up and form stable 'ridges' and 'piles' (Hansen and Yuen, 1988; Tackley, 1998, 2002; Table 1). The maximum height of the thermochemical bumps is the height of 'neutral buoyancy,' where local compositional buoyancy is just balanced by local thermal buoyancy (**Figure 54**; Kumagai et al., 2008). Upon reaching this level, the dome can, depending on its internal Rayleigh number, either stagnate or collapse. This generates very complicated morphologies (**Figure 54**) and time evolution (Kumagai et al., 2008).

7.03.9.2.4 Relevance of the experiments to the deep mantle: pressure dependence of B

The laboratory experiments span well the parameter space of the mantle, except in one respect: the depth dependence of the physical properties. This can be a problem since in the mantle, B could vary with pressure (see Oganov, volume 2), henceforth depth, because of two effects: (a) the thermal expansion coefficient α varies with depth (e.g., Hansen and Yuen, 2000; Montague and Kellogg, 2000; Tackley, 1998), and (b) the density contrast of compositional origin $\Delta\rho_\chi$ varies with depth since the two materials have different compressibilities (Samuel and Bercovici, 2006; Tan and Gurnis, 2005; Tan et al., 2011):

(a) The interface stability is determined by the ability of the thermal density contrast across the interface to overcome the compositional density contrast. The former should therefore be calculated using the value of a at the interface depth, and we do not expect B_c to vary significantly (see discussion in Le Bars and Davaille, 2004a). This is indeed confirmed by **Figure 53**, where both the numerical simulation results using a depth-dependent α and laboratory experiments collapse on the same diagram. In the mantle, α decreases with increasing depth (Oganov, volume 2), which, for a given compositional density contrast, has two effects on thermochemical convection. First, the regime of a given two-layer system will change with the location of the interface (Davaille, 1999b), from doming in the midmantle to stratified just above the CMB. Second, any dome starting deep in the mantle will gain buoyancy as it rises, which might enhance its velocity (Le Bars and Davaille, 2004b).

(b) Now, Si-enriched (i.e., denser) compositions would also induce a lower compressibility of the dense anomaly compared to the surrounding mantle, implying an increase of the compositional density excess with decreasing depth (Deschamps et al., 2011; Samuel and Bercovici, 2006; Tan and Gurnis, 2005; Tan et al., 2011). Therefore, although Si-enriched thermochemical domes could fully develop and rise toward the surface if $B < B_c$ initially, upon reaching the level of neutral buoyancy, they would stagnate and spread under it. If this level is significantly higher than the initial interface, large and well-formed plume heads could therefore stagnate at various depths of the lower mantle (Kumagai et al., 2008; Samuel and Bercovici, 2006). If the neutral buoyancy level is around the initial interface depth, tabular less viscous piles with steep vertical walls could form (Tan and Gurnis, 2005; Tan et al., 2011). This is the only way to generate less viscous active ridges or piles.

7.03.9.2.5 Stability of the continental lithosphere

Combining stability analysis and experiments, Cottrell et al. (2004) and Jaupart et al. (2007) studied the stability of the cooling continental lithosphere (see Parmentier, **Chapter 7.06**; and Jaupart, **Chapter 6.05**), a thin, compositionally lighter, but also more viscous, layer on top of a (quasi) infinite mantle. For this combination of transient cooling from above, small depth ratio a and large viscosity ratio γ, two types of oscillatory instabilities are obtained, depending on the value of B (Jaupart et al., 2007). In particular, for $0.275 < B < \sim 0.5$, oscillations are vertical and the lithosphere could grow by

entrainment of chemically denser material from the asthenosphere. These studies further suggest that the lithosphere could have developed in a state near that of instability with different thicknesses depending on its intrinsic buoyancy.

More recently, Fourel et al. (2013) investigated the morphology of these instabilities. Their pattern is made of linear upwellings and downwellings that converge radially toward the center of the continent in an outer region and an array of polygonal cells in a central region (**Figure 55**). The dynamics depends on three dimensionless numbers, the Rayleigh number for the unstable block Ra_{cont}, a buoyancy number that scales the intrinsic compositional density contrast to the thermal one B_{cont}, and the aspect ratio of the continental lithosphere h_{cont}/R_{cont}. In geologic conditions, such instabilities should manifest themselves as linear rifts striking at a right angle to the continent–ocean boundary – as the Cameroon line (Milelli et al., 2012) – and an array of domal uplifts, volcanic swells, and basins in the continental interior. Simple scaling laws for the dimensions and spacings of the convective structures give values in the 500–1000 km range, close to geologic examples. Moreover, the large intrinsic buoyancy of Archean lithospheric roots prevents this type of instability, which explains why the widespread volcanic activity that currently affects Western Africa is confined to post-Archean domains (Fourel et al., 2013).

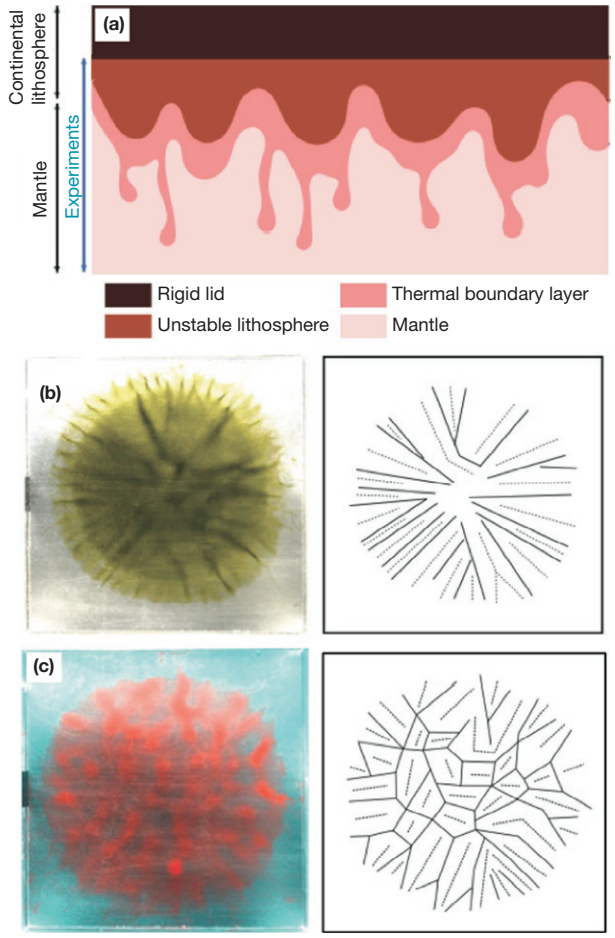


Figure 55 Instabilities of the continental lithosphere: (a) The unstable compositional lithosphere in light brown has a contorted base, with purely thermal small-scale convection in dark pink still present in the asthenosphere below (light pink). The two forms of convection are associated with different wavelengths and characteristic times. (b) Experiment with a small inner domain exhibiting well-marked radial structures ($Ra_{cont} = 41$, $B_{cont} = 0.23$, $h_{cont}/R_{cont} = 0.02$). (c) Experiment with a large inner domain showing the complicated polygonal inner pattern ($Ra_{cont} = 203$, $B_{cont} = 0.21$, $h_{cont}/R_{cont} = 0.048$). Reproduced from Fourel L, Milelli, L, Jaupart C, and Limare A (2013) Generation of continental rifts, basins, and swells by lithosphere instabilities. *Journal of Geophysical Research – Solid Earth* 118: 2169–9356.

7.03.9.2.6 Evolution of an initially stratified mantle and continental growth

The laboratory experiments always show an evolution from an initially stratified to a homogeneous fluid, either through continuous entrainment at the interface (see **Section 7.03.9.3.1**) or through thermochemical instabilities. Assuming that the creation of compositional heterogeneities through crustal formation is negligible, the timescales in the mantle would range from 100–300 My to Gy, depending on the initial buoyancy ratio and the amount of radioactive heating. Moreover, the sudden destabilization of a denser layer at the bottom of the mantle would create a succession of global thermochemical plume events, which would also accelerate briefly convective motions in the mantle and therefore subduction. This could lead to pulses in large igneous provinces (LIP) (via plumes) and granitoid (via subduction) formation, leading to crustal growth (Arndt and Davaille, 2013). This would explain well why the U–Pb ages of zircons from Precambrian granitoids and major rivers are grouped into a series of major peaks centered at about 2.7, 2.5, 2.1, 1.9, and 1.1 Ga (e.g., Condie et al., 2009): the first peak at 2.7 Gy would therefore correspond to the first destabilization of the initial bottom layer, while the following peaks would correspond to the subsequent doming events. Early episodic continental growth could therefore have been due to episodic convection in the mantle (e.g., Condie, 1998; Davies, 1995; Moorbath and Taylor, 1981; O'Neill et al., 2007; Reymer and Schubert., 1986; Stein and Hofmann, 1994) rather than to enhanced preservation through supercontinent assembly (i.e., Hawkesworth et al., 2009).

7.03.9.2.7 Diversity of mantle upwellings

As described by Ballmer et al. (this volume), hot spots and mantle upwellings present a large range of different signatures, which are difficult to reconcile with one single origin: there are probably different kinds of hot spots and upwellings in the mantle. According to Table 1 and **Figures 53** and **54**, this diversity could be explained by the coexistence of the different regimes of thermochemical convection in a compositionally heterogeneous mantle that convects today primarily on one layer (e.g., Davaille, 1999a,b; Davaille et al., 2005). According to **Section 7.03.7.3**, these thermochemical upwellings would develop primarily away from downwellings (e.g., Androvandi et al., 2011; Gonnermann et al., 2004; McNamara and Zhong, 2004a,b), that is, in two mantle 'boxes' separated by the circum-Pacific subduction zones (e.g., Davaille et al., 2002; McNamara and Zhong, 2005).

Then, long-lived hot spots could be due to plumes anchored on the topography of a dense ($B > 0.5$) basal layer, which could also develop piles and ridges. Scaling laws based

on laboratory experiments (Davaille, 1999a,b; Davaille et al., 2002; Jellinek and Manga, 2004) suggest that plumes anchored in D″ could develop for $\Delta\rho_\chi > 0.6\%$ and that they could survive hundreds of millions of years, depending on the spatial extent and magnitude of the anchoring chemical heterogeneity (Davaille et al., 2002; Gonnermann et al., 2002; Jellinek and Manga, 2002, 2004). Such plumes could therefore produce much longer-lived hot spots than the transient plumes observed experimentally in isochemical convection (**Figure 29**). However, plume and piles longevity does not necessarily imply complete spatial fixity: plumes and their anchors could be slowly advected by large-scale flow associated with strong downwellings such as subducting plates (Davaille et al., 2002; McNamara and Zhong, 2004a,b; Tan et al., 2002, 2011). In this framework, this implies that the relative inter-plume fixity over ~100 My is due to the relative steadiness of subduction zones.

On the other hand, 'superswells,' hot spot clusters, and large seismic slow anomalies in the deep mantle could be due to doming of a compositionally denser layer with $B < 0.5$. Then, according to eqn [34], thermochemical instabilities with temperature anomalies 300–500 °K could be produced at the base of the mantle by compositional density heterogeneities of 0.3–0.6%. If viscosity depends only on temperature, the instabilities would then be 10-3 000 times less viscous than the bulk mantle and should therefore have the form of cavity plumes. According to eqn [33], the thickness of the dense less viscous layer from which they come must be <200–700 km. The scaling laws of Le Bars and Davaille (2004a,b) predict velocities ~5–20 cm year^{-1}, radius ~500–1000 km, spacing 2000–3500 km, and cyclicity ~100–200 My for cavity plumes at the base of the mantle. Moreover, heat flux could vary by as much as 200% both laterally and spatially on those same timescales and length scales (Davaille et al., 2003; Le Bars and Davaille, 2004b; Namiki and Kurita, 2003). Note that the spacing predicted being much smaller than the typical size of a mantle 'box,' several instabilities should be observed in each box.

One fundamental question now remains: how might the large-scale circulation created by plate tectonics influence the different regimes described earlier, especially their time dependence? Can cyclic doming be maintained in such a flow? Experiments including compositional heterogeneities and plate-scale motions are still needed.

7.03.9.3 Entrainment and Mixing

7.03.9.3.1 Entrainment from an interface

The experiments of Olson (1984) on two-layer convection established that entrainment at low Reynolds number can indeed occur and that viscous stresses can substitute for inertial instabilities as a mixing mechanism. The entrainment occurs in two steps: first, the thermal heterogeneities at the interface induce circulations in the two layers; then, the viscous drag due to those convective motions becomes sufficient to overcome the negative buoyancy forces due to the stable chemical density gradient, and thin tendrils of material are entrained (**Figures 54(g)** and **56(a)**; Lister, 1989; Sleep, 1988). The entrainment rate and the filament thickness depend on the intensity of convection, its geometry, the viscosity ratio, and the buoyancy ratio: the more stable the fluid, the harder it is to entrain. A significant entrainment corresponds also to a cusp height greater than a critical value comparable to the thickness of the TBL above the interface (Jellinek and Manga, 2004). Scaling laws have been derived (Davaille, 1999a,b; Davaille et al., 2002; Jellinek and Manga, 2002, 2004) that explain the data well (e.g., **Figure 56(b)**).

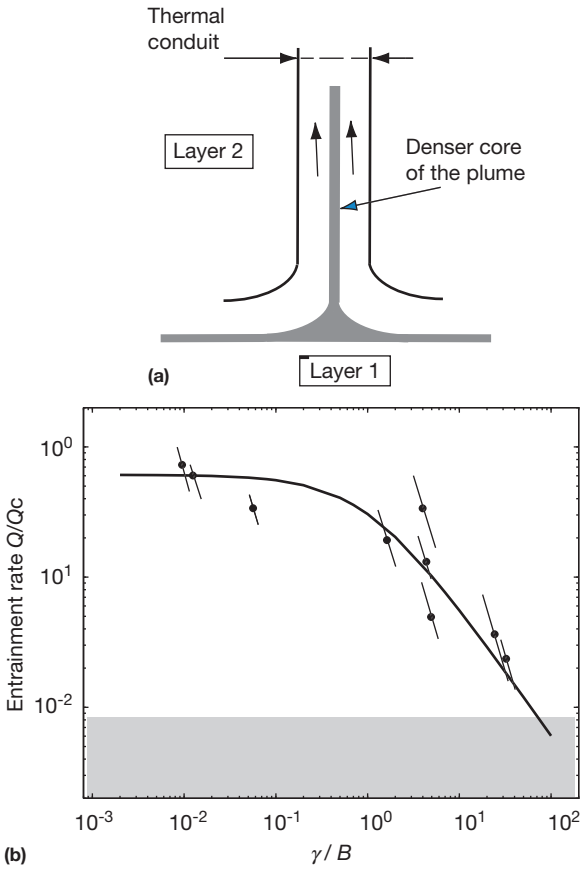


Figure 56 (a) Sketch of a thermochemical plume. (b) Volumetric entrainment rate of a thermochemical plume divided by the convective scale as a function of γ/B. The circles represent the experimental measurements, while the solid line stands for the theoretical scaling. The shaded area shows the domain where salt diffusion becomes important. Reproduced from Davaille A, Girard F, and Le Bars M (2002) How to anchor plumes in a convecting mantle? *Earth and Planetary Science Letters* 203: 62634.

7.03.9.3.2 Stirring

The mixing pattern in thermochemical convection (**Figure 57**) is very complex and still poorly characterized since convection creates compositional heterogeneities with two different typical sizes and topologies: (a) thin filaments generated by mechanical entrainment across the interface, leading to the development of marble cake structures in both reservoirs, even though they can remain dynamically separated for a long time, and (b) domes, and blobs encapsulated within them, which are generated by instabilities.

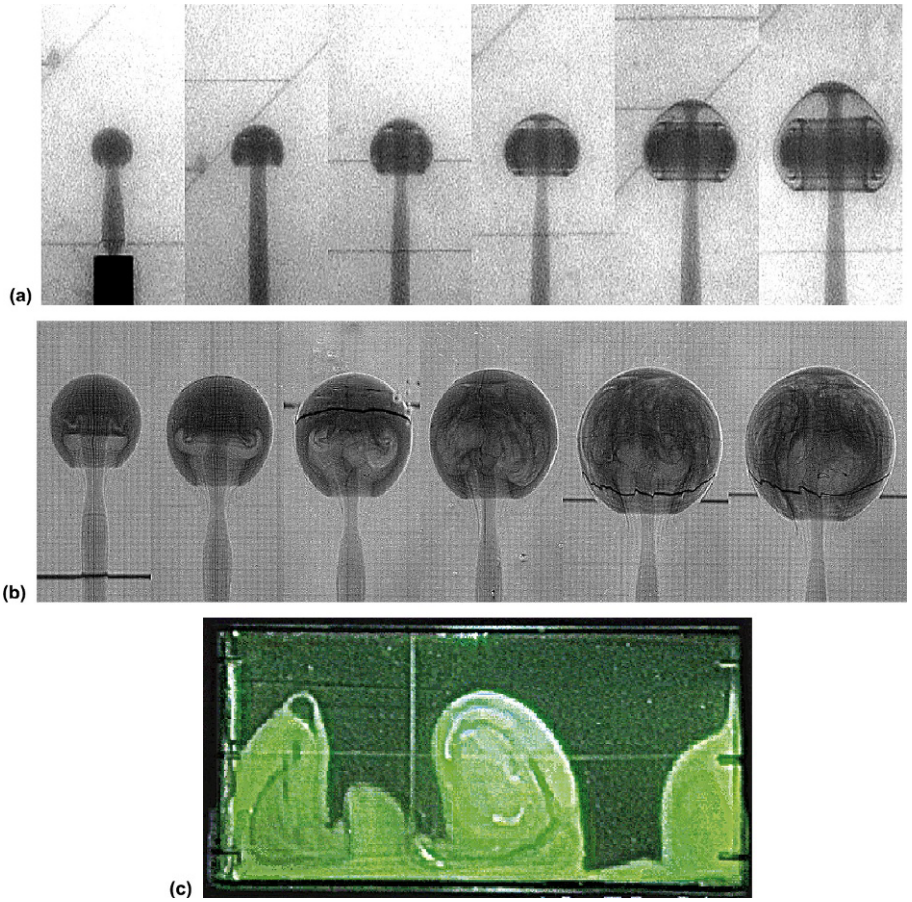


Figure 57 Entrainment in compositional plumes (a and b) and thermochemical domes (c) by viscous coupling. (a) Low viscosity ratio: stirring within the plume head follows a vortex ring structure. (b) High-viscosity ratio: stirring within the head is chaotic (reproduced from Kumagai I (2002) On the anatomy of mantle plumes: Effect of the viscosity ratio on entrainment and stirring. *Earth and Planetary Science Letters* 198: 211224). (c) Mixing pattern in whole-layer thermochemical convection. The lower layer, initially denser and more viscous, is dyed with fluorescein. The experimental tank is illuminated by a laser sheet. After three cycles, the two layers are still not mixed, but one can distinguish filaments and blobs of dark, light fluid within the viscous domes. The thin filaments were entrained by viscous coupling, while the blobs were encapsulated when the viscous dome collapsed upon cooling. Since the blobs are lighter, they are bursting up through the viscous fluid as Rayleigh–Taylor instabilities. ($B = 0.30$; $Ra = 2.7 \times 10^7$; $a = 0.3$; $\gamma = 175$). Reproduced from Davaille A, Le Bars M, and Carbonne C (2003) Thermal convection in a heterogeneous mantle. *Comptes rendus de l'Académie des Sciences Gosciences* 335: 141156.

7.03.9.3.3 Experimental limitation: species diffusion

The geodynamic irrelevance of surface tension effects requires the use of miscible fluids in the laboratory. Therefore, chemical diffusion becomes a problem when the ratio of chemical to thermal diffusivity (the Lewis number) $Le = D/\kappa$ is too small. In the mantle, this ratio is $>10^9$. In the laboratory experiments reported here, it was always >1000. Even then, chemical diffusion becomes important when convection is slow (cf. Section 7.03.9.1), and the mechanism of entrainment across the interface then switches to advection-enhanced diffusion (see the discussion in Davaille, 1999a, and gray area in Figure 56(b)). Another problem arises when three-component solutions are used (e.g., water + salt + Natrosol) in which the two dissolved species (e.g., salt and Natrosol) do not diffuse at the same speed. Then, if the more viscous layer is initially the upper layer (i.e., the layer with no salt is the layer with the most Natrosol), salt–Natrosol fingers develop, even before the onset of thermal convection. This phenomenon, well known in thermohaline convection, changes drastically the physics and the entrainment rate at the interface. This led Davaille (1999a,b) to discard those experiments. This could also explain the results of Namiki (2003), who observed, using Natrosol solutions, that the interfacial entrainment rate when the more viscous layer was on the top was different than when it was at the bottom.

7.03.9.4 Interaction of a Plume with a Density and/or Viscosity Interface

This situation is of interest for the mantle since the 660 km depth seismic discontinuity corresponds to the phase transition from the spinel phase at low pressure to perovskite at higher pressure (see Oganov, Chapter 2.08). Although this endothermic phase transition probably cannot stratify mantle convection, it is sufficient to delay flow (see Ricard, Chapter 7.02).

Moreover, the perovskite phase is also 3–100 times stiffer than the upper mantle phases.

Laboratory experiments coupled to numerical modeling have shown that a viscosity jump alone is sufficient to modify the shape of an upwelling compositional plume, the head of which becomes elongated as it crosses the discontinuity from the stiffer lower mantle to the less viscous upper mantle (Manga et al., 1993). For sufficiently large viscosity contrast, the plume head can even accelerate so much in the upper mantle that it becomes disconnected from the plume tail, which aggregates at the interface and forms a second plume head (Bercovici and Mahoney, 1994). If the plume is compositionally heterogeneous, segregation can occur at the interface, and the two head events then have different geochemical signatures (Kumagai and Kurita, 2000). These phenomena would explain well the emplacement timing and the geochemical data at the Ontong Java Plateau (Bercovici and Mahoney, 1994; Kumagai and Kurita, 2000).

Kumagai et al. (2007) extended the previous studies to thermal plumes and investigated the interaction between a lower layer thermal plume and an inner interface. The interaction mode depends on the local buoyancy number (B_L, the ratio of the stabilizing chemical buoyancy to the plume thermal buoyancy at the interface), Ra, and γ. For $B_L < 0.6$, the pass-through mode develops, whereby a large volume of the lower material rises through the upper layer and reaches the top surface, since the plume head has a large thermal buoyancy compared to the stabilizing density contrast between the two layers. When $B_L > 0.6$, the rebirth mode occurs, where the thermal plume ponds and spreads under the chemical boundary and secondary thermal plumes are generated from the interface. Depending on the magnitude of B_L, these plumes can entrain a significant amount of lower layer material upward by viscous coupling.

7.03.10 Mid-Ocean Ridges and Wax Tectonics

The study of ridges and accretion is challenging because it involves processes related to both brittle fracture and ductile flow, which is quite difficult to study numerically. For example, despite significant steps in the past years (for a recent review, see Gerya, 2012), current numerical simulations still cannot reproduce the coupled fluid–solid deformation processes responsible for microplates at mid-ocean ridges. On the other hand, published results from a wax analogue model yielded the first observations of overlapping spreading centers (OSCs), morphological precursors to microplates, before their discovery on the seafloor (Oldenburg and Brune, 1972). Two main questions have been tackled with laboratory experiments: the ability of a buoyant mantle upwelling to break the crust and the morphology of ridges.

7.03.10.1 Can a Buoyant Mantle Upwelling Generate Sufficient Stress to Rupture a Brittle Crust?

Using surface layers of various brittle and viscoelastic materials, Ramberg and Sjostrom (1973) studied whether a relatively stiff crustal layer could break in tension above a buoyant diapir and become displaced laterally in a manner simulating the breakup of Pangaea and the wandering continents. For better visualization of the deformation, the material was initially dyed in different colored layers. The model was then centrifuged for several minutes at accelerations up to 3000g. All experiments show the breakup of the brittle layer by the diapir head. However, the material rheology was not sufficiently well characterized to permit a quantitative study or to investigate the morphology of the ridges. With the same kind of setup, Tentler (2003a,b, 2007) investigated the morphology, propagation, and interaction of the crustal fractures developing in the first steps of ridge initiation above an upwelling. Acocella (2008) and Tentler and Acocella (2010) further studied the influence of the initial configuration of precut ridge segments on their subsequent interaction. However, only the brittle crust was considered, although mantle deformation needs to become localized also in the ductile part of the lithosphere.

7.03.10.2 Morphology of Ridges

Ridge morphology has been studied using paraffin wax as an analogue mantle material: solid wax simulates brittle mantle and molten material simulates the region that deforms as a viscous fluid (Oldenburg and Brune, 1972, 1975; Ragnarsson et al., 1996). The molten wax was frozen at the surface by a flow of cold air. Then, the solid crust was pulled apart with constant velocity and a rift was formed separating the crust into two solid plates (Figure 58). In this case, the solid wax thickness increases as the square root of the distance to the ridge (Oldenburg and Brune, 1975), in agreement with theory (Parker and Oldenburg, 1973).

Oldenburg and Brune (1972) first observed that a straight rift initially perpendicular to the pulling direction evolved into

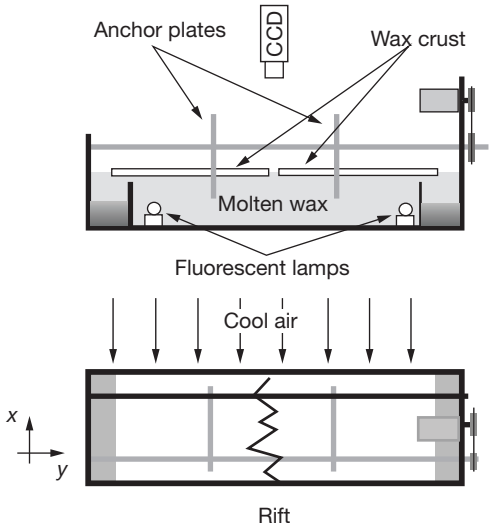


Figure 58 Experimental setup. The tank is heated from below at constant temperature and cooled from above by a constant flow of air. Before each run, a wax layer of wax is allowed to grow on the surface until thermal equilibrium is reached. Two skimmers embedded in the solid wax are attached to a threaded rod, which is driven by a stepping motor. The rift is initiated with a straight cut through the wax, perpendicular to the spreading direction. Divergence at this cut causes liquid wax to rise into the rift and solidify. Illumination from below permits to image the plate thickness at the rift from above using a video camera. Reproduced from Ragnarsson R, Ford JL, Santangelo CD, and Bodenschatz E (1996) Rifts in spreading wax layers. *Physical Review Letters* 76: 3456–3459.

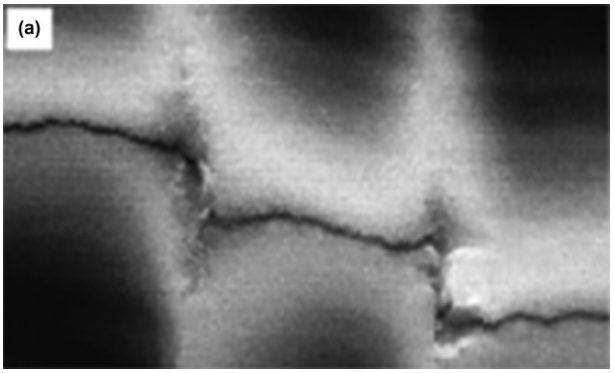


Figure 59 Transmission image of transform faults, obtained in Shell Wax Callista 158 and with a spreading rate of 78 mm s^{-1}. Image size is 9 mm. Reproduced from Bodenschatz E, Gemelke N, Carr J, and Ragnarsson R (1997) Rifts in spreading wax layers: An analogy to the mid-ocean rift formation. *Symposium on Localization Phenomena and Dynamics of Brittle and Granular System*. Columbia University.

a pattern consisting of straight segments interrupted by faults orthogonal to the rift and parallel to the pulling direction (Figure 59). The ability of a wax to generate orthogonal transform faults depends on the ratio of the shear strength of its solid phase to the resistive stresses acting along the transform fault: if the latter exceed the former, a breakup of the solid near the transform fault wall should occur (Oldenburg and Brune, 1975).

Ragnarsson et al. (1996) and Bodenschatz et al. (1997) studied the phase diagram in more detail, investigating the rift structure as a function of the pulling speed for two different waxes. The regime diagram was found to depend critically on the rheology of the wax:

- In the first case (Shell Wax 120), the solid layer could be divided in two regions because the paraffin wax undergoes a solid–solid phase transition: a colder phase, hard and brittle, where strain is mostly accommodated by crack formation during extension; and a warmer ductile phase, where strain is accommodated primarily by viscous flow. This wax gave birth to the different regimes seen on Figure 60 (Ragnarsson et al., 1996), but the orthogonal transform faults observed by Oldenburg and Brune (1972, 1975) were never observed. At slow spreading rates, the rift, initially perpendicular to the spreading direction, was stable (Figure 60(a)). Above a critical spreading rate, a 'spiky' rift developed with fracture zones almost parallel to the spreading direction (Figure 60 (b)). At yet higher spreading rates, a second transition from the spiky rift to a zigzag pattern occurred (Figure 60(c)). With further increase of the spreading rate, the zigzags steepened (Figure 60(d)). In Figure 60(a) and 60(b), the rift was frozen, while in Figure 60(c) and 60(d), molten fluid reached the surface. In the zigzag regime, the angle of the interface with respect to the spreading direction can be described by a simple geometric model, since the problem is dominated by only two velocities, the externally given spreading rate (pulling speed) and the internally selected maximal growth velocity of the solidifying fronts. When the spreading rate exceeds the solidification speed, the interface has to grow at an angle to keep up. The experimental data on the angles agree well with this model.

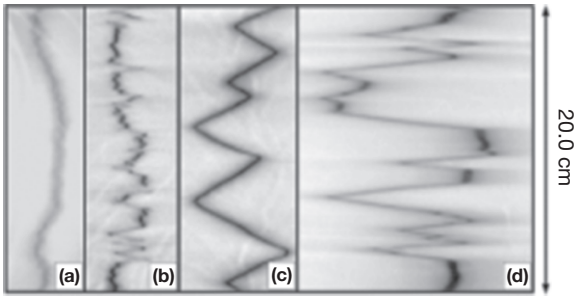


Figure 60 Different regimes as the spreading velocity increases. Reproduced from Ragnarsson R, Ford JL, Santangelo CD, and Bodenschatz E (1996) Rifts in spreading wax layers. *Physical Review Letters* 76: 3456–3459.

- In the second case (Shell Wax Callista 158), the solid phase is more brittle (Bodenschatz et al., 1997; Oblath et al., 2004). Three distinct morphological regimes were observed. At slow spreading rates, a straight rift is stable and forms a topographic low. At moderate rates, it becomes unstable and OSCs and microplates form, evolve, and die on the ridge, which has little or no relief. At high spreading rates, the microplates lose their internal rigidity and become transform faults at the ridge (Figure 59), as previously described by Oldenburg and Brune (1972, 1975). The ridge corresponds to a topographic high (Bodenschatz et al., 1997; Oblath et al., 2004). Katz et al. (2005) focused on the formation of microplates (Figure 61). In wax, like on Earth, they originate from OSCs. The latter nucleate predominantly on sections of obliquely spreading rift, where the rift normal is about 45° from the spreading direction.

These experiments are very interesting for pattern formation and show striking geometric similarities with what is observed on the ocean floor. However, their relevance to mid-ocean ridges systems is still questionable for two reasons: First, dynamic similarity is not obtained, since the viscosity contrast between the solid and the liquid wax is much greater than the viscosity contrast across the solid lithosphere. Experiments involving wax are therefore probably more relevant to the dynamics of ice satellites (e.g., Manga and Sinton, 2004) or of lava lakes. Second, the physics of the wax experiments reported earlier is far from being understood. For example, the influence of latent heat effects on the dynamics of the wax systems remains unknown. To apply their results to the Earth's mid-ocean ridges system, Oldenburg and Brune (1972, 1975) neglected these effects and proposed a semiquantitative theory based on shrinkage and the mechanical properties of wax. The more recent experiments, although they very nicely showed the diversity of rift morphologies, were not able to relate them quantitatively to the wax cooling, nor to its rheology. So a quantitative understanding of lithosphere rifting remains to be achieved and will require a better knowledge of the rheology.

7.03.11 Subduction-Related Experiments

The subduction of oceanic lithosphere is the most prominent signature of thermal convection in the Earth's mantle. Subducting plates constitute the cold downwellings of mantle

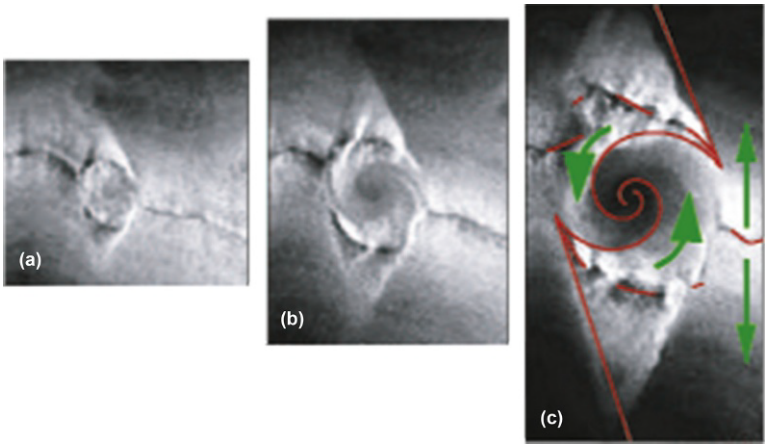


Figure 61 Time series of images showing a growing microplate. (a) 15 s, (b) 30 s, (c) 60 s. Green arrows show the direction of spreading of the main plates and the direction of microplate rotation. In red, the spreading ridge and the pseudofaults pair. The inner ones were generated with the kinematic model of Schouten et al. (1993). Reproduced from Katz et al. (2005).

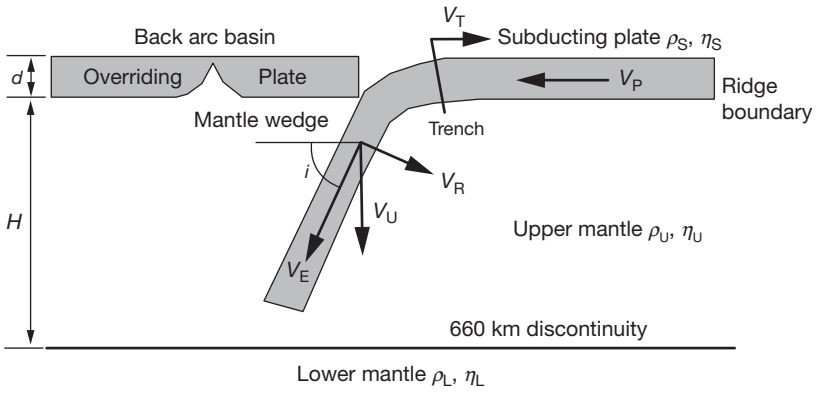


Figure 62 Sketch of a subduction zone with the main features studied in the laboratory. V_P: surface plate velocity; V_T: retrograde trench velocity; V_R: retrograde slab velocity; V_E: down-dip velocity; V_U: sinking velocity; i, dip angle; d, slab thickness; H, upper layer thickness.

convection, which sink into the mantle because of their negative buoyancy. Subduction involves a wide variety of physical and chemical processes: earthquakes, volcanism (dehydration, melting, and melting migration), phase transformations, thermal effects, mantle circulation, and plate motion. They are reviewed in detail by King in this volume. We focus here on the experimental work that has been restricted to the study of the mantle-scale dynamics of subduction, namely, the relations between motion of subducted slabs, mantle flow, and back-arc spreading (Figure 62). Laboratory models of these phenomena have three advantages: They are inherently 3-D; they do not suffer from limited computational resolution of the temperature and velocity fields; and it is easy to work with strong viscosity variations.

7.03.11.1 Ingredients for Subduction

It has been long recognized that cold plates are not produced by isoviscous convection and that the strong temperature dependence of mantle viscosity is required (e.g., Torrance and Turcotte, 1971), as we showed in Sections 7.03.5–7.03.6. Turner (1973) illustrated this idea using glycerine in a tank cooled from above, in which convective overturn was forced by a stream of bubbles of cold CO_2 released from dry ice at the bottom of the tank. Surface cooling then produced a thin viscous sheet that was driven away from regions of upward convection and then plunged steeply into the less viscous interior, maintaining its identity as it did so. Jacoby (1970, 1973, 1976) and Jacoby and Schmeling (1982) studied the gravitational sinking of cold viscous high-density lithosphere into asthenosphere. Jacoby first studied the mode of sinking of a heavy elastic rubber sheet into water under isothermal conditions without active 'mantle' convection (Jacoby, 1973). Later, he used the same paraffin as Oldenburg and Brune (1975) to include the effects of thermal convection. The paraffin was heated from below above its melting point and cooled from above below its melting point. It therefore formed a thin solid dense skin on top of the fluid layer. However, the skin could not be generated or melted at the same rate as that of the sinking, preventing real plate tectonics behavior (Jacoby, 1976). Moreover, surface tension prevented even thick skins from sinking by themselves. However, sinking could be

initiated by loading the edge of the skin or by wetting its surface with molten paraffin (Figure 63). The heavy skin first bends with a radius of curvature, which depends on the skin thickness, and then accelerates down the trench as the portion within the mantle increases. Upon interacting with the tank bottom, the skin can fold and the trench can even roll back in the oceanward direction (Figure 63). Convection cells caused by localized heating at the bottom of the tank do not perturb significantly the slabs evolution. Although these pioneering experiments were only 'semiquantitative' (Jacoby, 1976), they already reproduced most features of subduction (down-dip motion and roll back, 3-D, time dependence, folding on the box bottom, etc.), highlighting the difficulties that quantitative experiments have to overcome (subduction initiation, surface tension effects, etc.).

Kincaid and Olson (1987) designed experiments to further study the lateral migration of slabs and their penetration through the transition zone (Figure 62). Cold, negatively buoyant molded slabs of concentrated sucrose solution were introduced into a more dilute, two-layer sucrose solution representing the upper and lower mantles. The transition zone was modeled by a step increase in both density and viscosity. The initial setup consisted of two horizontal plates separated by a trench gap, with one plate attached to a dipping slab, simulating a developing subduction zone with an overriding plate (Figure 64). The trailing end ('ridge') of the slab was either locked ($V_P = 0$) or free to move along with the plate. Besides the Prandtl number, important dimensionless numbers are the ratio of slab thickness to mantle depth δ/H and the ratios η_S/η_U and η_s/η_L of the slab viscosity to those of the upper and lower mantles. It was necessary to have the $\eta s/\eta_L > 10^3$ to maintain a tabular shape during subduction. No velocity is imposed and the slab sinks under its own weight. This phenomenon can be described by an effective Rayleigh number, defined in terms of the slab/upper layer density contrast $\rho_S - \rho_U$:

$$Ra_S = \frac{(\rho_S - \rho_U) g H^3}{\kappa \eta_U} \quad [35]$$

In the experiments, $Ra_S \sim 1.2 \times 10^5$. On the other hand, the style of slab penetration through the density discontinuity depends on the ratio of the slab buoyancy in the upper and the lower mantle:

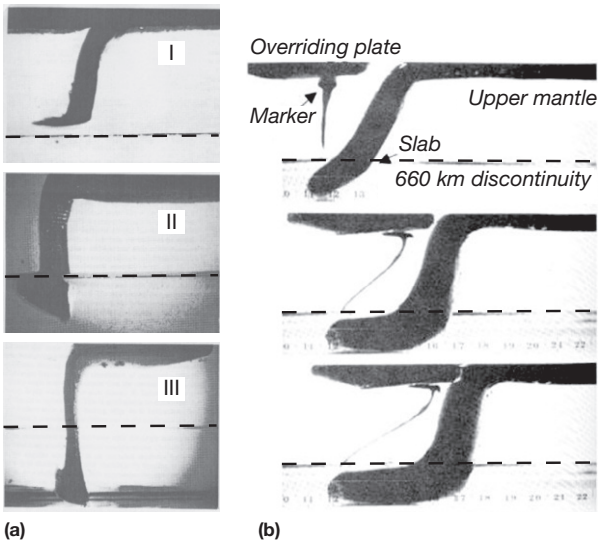


Figure 64 (a) Styles of slab penetration through a density discontinuity: I. Slab deflection with $B_S \sim -0.2$; II. Partial slab penetration and formation of a root beneath the interface with $B_S \sim 0$; III. Complete slab penetration for $B_S > 0.5$. (b) Trench migration and retrograde subduction with the fixed ridge boundary condition. The dye streak indicates the flow pattern in the back-arc edge. Reproduced from Kincaid C and Olson P (1987) An experimental study of subduction and slab migration. *Journal of Geophysical Research* 92: 13832–13840.

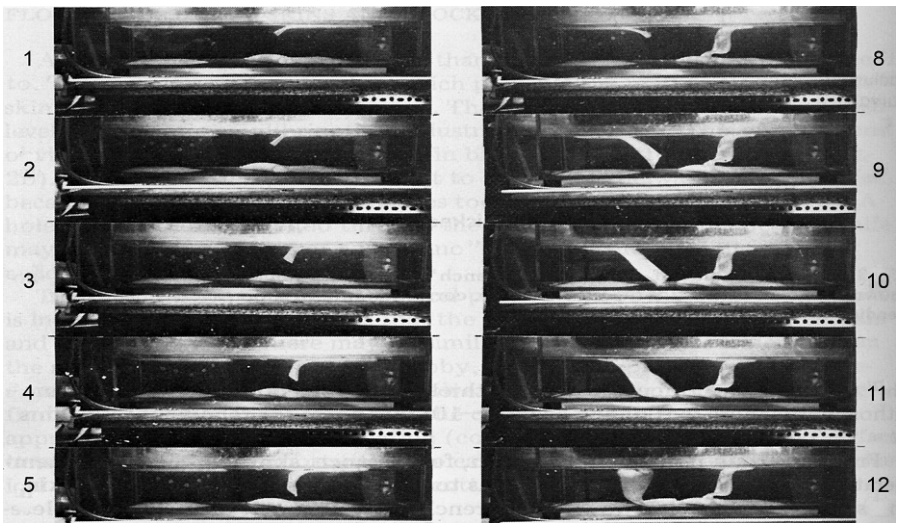


Figure 63 Series of photographs of two sinking paraffin skins, taken at ~3 s intervals with a longer interval between 5 and 8. Changes in slab dip, retrograde motion, and folding on the bottom boundary are clearly seen. Reproduced from Jacoby WR (1976) Paraffin model experiment of plate tectonics. *Tectonophysics* 35: 103–113.

$$B_S = \frac{\rho_S - \rho_L}{\rho_S - \rho_U} \qquad [36]$$

Homogeneous fluids correspond to $B_S = 1.0$, while negative values correspond to strong stratification. In agreement with the numerical study by Christensen and Yuen (1984), Kincaid and Olson (1987) found three modes of slab penetration (**Figure 64(a)**). Regime I, in which $B_S \leq -0.2$, corresponds to strong stratification and virtually no slab penetration. In the intermediate regime II ($-0.2 \leq B_S \leq 0.5$), there is partial slab penetration and formation of a root beneath the interface. When $B_S > 0.5$, the stratification is weak enough to permit the slab to sink into the lower layer with minor deformation (regime III). Retrograde slab motion and trench migration occurred in nearly every case, being greatest in regime I (**Figure 64(b)**), episodic in regime II, and nearly absent in regime III, especially when η_L/η_U is around 1. Moreover, the episodicity of trench rollback is related to the interaction of the slab with the density stratification and is greatest in the case of a fixed ridge. The complexity of the time-dependent slab motion, both down-dip and retrograde, and its interaction with any bottom boundary made it difficult to derive scaling laws to describe subduction. Work in the last 25 years has aimed at obtaining those scalings, using experiments where more parameters were controlled.

7.03.11.2 The Surface Story and the Initiation of Subduction

Subduction initiates within the lithosphere and produces large-scale tectonic features. The first analogue modeling of tectonic processes probably dates back to 1815 when Sir James Hall attempted to model folds in geologic strata, using layered beds of clay and clothes. Numerous experiments have been performed since to reproduce lithosphere and crustal dynamics. It is beyond the scope of this chapter to review them all (for recent reviews, see Becker and Faccenna, 2009; Billen, 2008; Schellart, 2002). We shall describe here only studies relevant to the nucleation and evolution of subduction.

Lithospheric processes necessitate taking into account the complex rheology of the lithosphere. Shemenda (1981; 1992; 1993) considered the subduction of an elastoplastic lithosphere (a mixture of powder and hydrocarbons) into a low-viscosity asthenosphere (water), driven by both a horizontal gravitational sinking (i.e., plate/asthenosphere density contrast) and a horizontal compressional force driven by a piston. However, subduction did not initiate without the presence of preexisting discontinuities. In this case, the horizontal compression produced buckling instabilities and then localization of deformation, leading to failure and eventually subduction of the lithosphere. Trench rollback was then accompanied by back-arc opening along preexisting faults. The lithosphere has also been modeled as a succession of brittle and ductile layers of different strengths and rheologies (e.g., Davy and Cobbold, 1991). Following this approach, Pinet and Cobbold (1992) and Pubellier and Cobbold (1996) studied the consequences of oblique subduction, that is, the partitioning between down-dip motion and transverse motion along faults. They used a sand mixture to model the brittle upper crust, silicone putty for the ductile lower crust/upper mantle, and glucose syrup for the asthenosphere.

Faccenna et al. (1996, 1999) and Becker et al. (1999) used the same layered system to study the behavior of an ocean–continent plate system subjected to compressional strain over geologic timescales. Compressional stress was achieved by displacing a piston at constant velocity perpendicular to the plate margin (**Figure 65**). The coupled interface between oceanic and continental plates is found to resist a rapid surge of compressional stress, and the shortening is accommodated by undulations within the ocean plate (Martinod and Davy, 1994). But if the system evolves under a low compressive strain rate (slow ridge push), the oceanic plate becomes unstable to RTI instability and subduction develops (Faccenna et al., 1999). The trench is then first localized at the ocean–continent boundary because of the lateral heterogeneity there. The dynamics of the system can therefore be controlled by the 'buoyancy number' F, the ratio of driving buoyant to resisting viscous forces (e.g., Houseman and Gubbins, 1997):

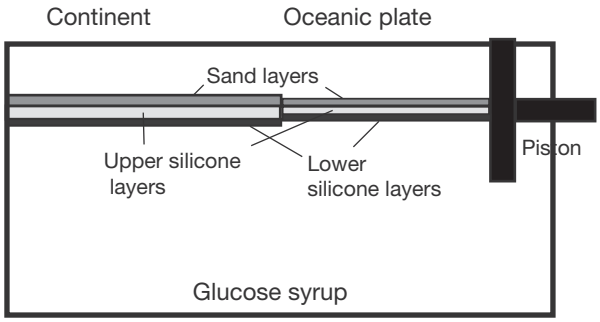


Figure 65 Four-layer setup to study the initiation of subduction. Reproduced from Faccenna C, Giardini D, Davy P, and Argentieri A (1999) Initiation of subduction at Atlantic-type margins: Insights from laboratory experiments. *Journal of Geophysical Research* 104: 2749–2766.

$$F(\rho_S - \rho_U, \delta_S, \eta_S, U) = \frac{(\rho_S - \rho_U) g \delta_S^2}{\eta_S U} = \frac{\sigma_{RT}}{U/\delta_S} \qquad [37]$$

where U is the imposed horizontal velocity and σ_{RT} is the RTI growth rate for a more viscous slab (see **Section 7.03.4**). If $F(\rho_S - \rho_U, \delta_S, \eta_S, U) < 1$, the RTI is inhibited and the oceanic plate is deformed by folding. Subduction initiates if $F(\delta_S) > 1$, that is, if the negative buoyancy force of the oceanic plate exceeds the viscous resisting force of the model lithosphere. Faccenna et al. (1999) further showed that in their setup, the passive margin behavior is not sensitive to the high shear strength of the brittle layer but only to the resistance of the ductile layer: subduction initiation is essentially a viscous fluid process. Since its growth rate σ_{RT} depends on the inverse of the slab viscosity (which is high), it is a very slow process. However, once the developed slab sinks into the less viscous mantle, its style and dynamics are predominantly governed by a buoyancy number $F(\rho_S - \rho_U, H, \eta_U, U)$ based on the viscosity and the thickness of the upper mantle (Becker et al., 1999). The velocity, angle of subduction, and the rate of trench rollback are found to be strongly time-dependent and to increase exponentially over tens of millions years before the slab interacts with the 660 km discontinuity.

More recently, two independent sets of experiments showed that low-angle subduction could also be generated by the lateral density differences and topography contrasts existing between oceanic and continental domains in passive margins

(Goren et al., 2008; Lévy and Jaupart, 2012). It allows the continental crust to spread over the oceanic basement, so that the oceanic plate bends under the extra load. This mechanism requires at least 100–150 My to initiate subduction and a continental crust soft enough.

7.03.11.3 Mantle Flow Induced by a Cold Rigid Slab

The use of rigid slabs enables one to maintain constant along the slab the dip angle i and the slab material velocity V_E and to control easily both V_E and the velocity V_R of retrograde motion by varying either the trench velocity V_T or the time variation of the dip angle di/dt (Figure 62). In this setup, all plate motions are kinematically determined. They can be scaled to the mantle through the Péclet number $Pe = v_E \delta_s / \kappa$, which compares advective heat transport and conductive transport. For mantle subduction, $10 \leq Pe \leq 400$. This means that the slab will have reached 660 km depth before having lost much heat.

7.03.11.3.1 Generation of seismic anisotropy

Motivated by the growing body of evidence that trench rollback could be associated with a particular pattern of seismic anisotropy (see Montagner, volume 1; and Mainprice, volume 2), whereby the a-axis of olivine crystals would align parallel to the trench oceanward of the slab and turn around its edges (e.g., Russo and Silver, 1994), Buttles and Olson (1998) used small cylinders (whiskers) suspended in a viscous fluid as an analogue to olivine crystals and studied their orientation in vertical and horizontal cross sections in the vicinity of a subducting slab. The slab was simulated by a rigid Plexiglas plate whose dip angle, down-dip motion, and rollback were controlled independently (Figure 66), and the mantle was modeled with Newtonian corn syrup. The whole system was isothermal. They show that trench-parallel olivine a-axis orientation in the seaward-side mantle is indeed controlled by the amount of slab rollback and that orientations in the mantle wedge depend upon slab dip angle. However, when the rollback component is weak, the anisotropy distribution is more complex and depends strongly on the initial orientation of the markers and of the time evolution of subduction (Druken et al., 2011).

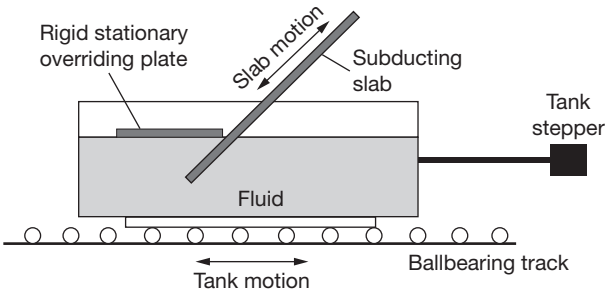


Figure 66 Typical experimental setup allowing to control independently the trench velocity and the slab motion (velocity and dip angle). The subducting slab can be either a rigid plate mounted on a stepper (e.g., Buttles and Olson, 1998; Kincaid and Griffiths, 2003, 2004) or a viscous fluid extruded at a constant rate from a thermostated reservoir (e.g., Griffiths et al., 1995; Guillou-Frottier et al., 1995).

7.03.11.3.2 Mantle flow and thermal evolution of the slab

Kincaid and Griffiths (2003, 2004) used the same kind of setup to study the variability of flow and temperature in subduction zones. The slab was made of a composite laminate cooled to a prescribed temperature before the experiment and contained temperature sensors to monitor its thermal evolution as it is forced into an isothermal tank of hotter glucose syrup. The TBLs developing in the fluid were visualized by a Moiré technique (Figure 67(a)) and the flow field measured using tiny air bubbles and PIV (Figure 67(b)). The experiments Pe ranged between 80 and 450. The mantle return flow induced by pure longitudinal sinking creates two cells, one oceanward and one in the wedge (Figure 67(b)), where the velocity can reach 40% of the slab down-dip speed. Rollback subduction induces flow both around and beneath the sinking slab (Figure 68), with larger velocities in the wedge and flow focused toward the center of the plate. These 3-D flows strongly influence the thermal evolution and structure of the plate: they speed up slab reheating. Moreover, the highest slab reheating occurs along the slab centerline when there is rollback, while it occurs along the edges for pure longitudinal sinking.

7.03.11.4 Coupled Mantle–Slab Fluid System

These experiments consider the interactions between the mantle and a mature, purely viscous slab. The subduction is induced either by extruding a viscous slab with constant velocity and dip angle (e.g., Figure 66) or by bending the tip of a horizontal slab and forcing it into the mantle until the slab sinks under its own weight (e.g., Figure 62).

7.03.11.4.1 Fixed rollback velocity

These experiments were designed to further investigate the interaction of a sinking slab with a density and viscosity interface, mimicking the discontinuity at 660 km depth between the upper and lower mantles (Kincaid and Olson, 1987). The slab velocity and initial dip angle were this time imposed. Griffiths and Turner (1988) injected tabular slabs and cylindrical plumes vertically ($V_T = 0$) onto a fluid interface under isothermal conditions and determined the conditions under which folding of the slab or plume occurred. Mixtures of water and golden syrup were used to control the viscosity and density of the slab and the mantle layers. As in the experiments of folding on a solid surface (Cruickshank and Munson, 1981; Figure 64(a)), they found that folding occurs in a compressed slab when the ratio of its length to its width exceeds a critical value (around 6). They also determined the amount of fluid entrained within the folds in the lower layer and when this entrainment could sufficiently reduce the buoyancy of the slab to prevent its sinking further in the lower mantle. Griffiths et al. (1995) and Guillou-Frottier et al. (1995) extended these results to retreating slabs (V_E and V_T constant). Griffiths et al. (1995) used again isothermal mixtures of golden syrup and water so that their slab was between 1.5 and 50 times more viscous than their lower mantle. Guillou-Frottier et al. (1995) used corn syrup and extruded a very cold and viscous slab into their two-layer mantle, so that $Pe \sim 30$–150 and the slab was 10^2–10^4 more viscous than the lower mantle. Both sets of experiments highlighted the essential role of trench rollback, in addition to B_S and η_l/η_u, in the slab penetration

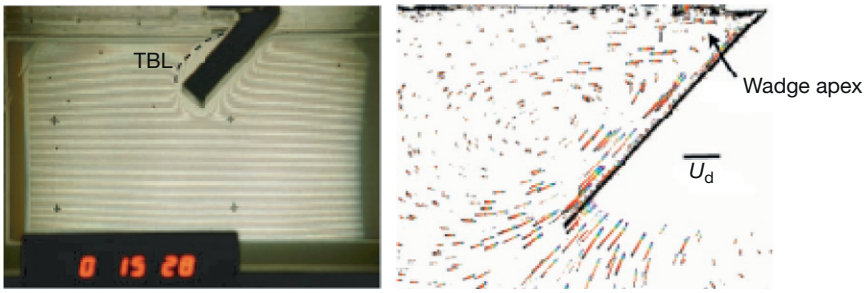


Figure 67 Rigid slab experiments. (a and,b) A cold rigid slab is introduced at constant velocity in the fluid, with no trench rollback. (a) The cold TBL developing around the slab is visualized with a Moiré technique. (b) Corner flow motions in the mantle wedge visualized with streaks. Reproduced from Kincaid C and Griffiths RW (2004) Variability in flow and temperatures within mantle subduction zones. *Geochemistry, Geophysics, Geosystems* 5: Q06002. http://dx.doi.org/10.1029/2003GC000666.

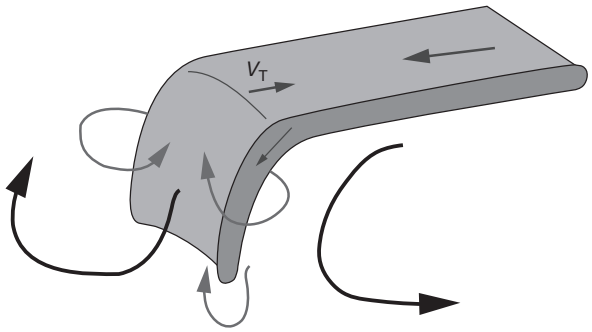


Figure 68 Schematic sketch of subduction and slab rollback-induced flow. Down-dip motion induces two poloidal flow cells on each side of the slab (black arrows), while slab rollback expels oceanward mantle material into the wedge, creating three toroidal flow cells, two around the lateral slab edges, and one around the tip of the slab when the slab has not reached yet the bottom of the box (or the density stratification).

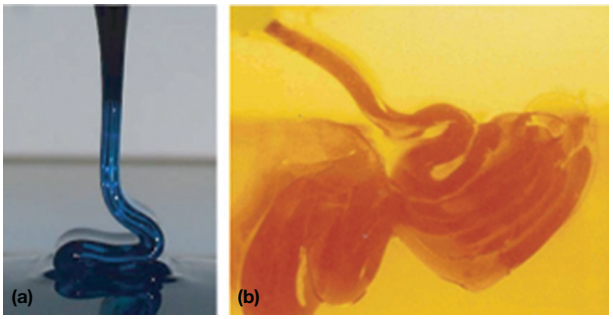


Figure 69 (a) A sheet of viscous fluid poured on a surface can develop folds and form a pile (reproduced from Ribe (2003)). (b) The same folding and piling behavior is observed when a viscous slab sinking in another viscous fluid encounters a viscosity and/or density interface. Here, the slab viscosity is 1000 times that of the lower layer (reproduced from Guillou-Frottier L, Buttles J, and Olson P (1995) Laboratory experiments on structure of subducted lithosphere. *Earth and Planetary Science Letters* 133: 1934).

regime. Increasing trench rollback enhances slab flattening on the interface (Figures 64 and 70). Increased resistance in the lower mantle also promotes folds and piles (Figure 69(b)), which can either stagnate in the lower mantle or sink to the bottom (Guillou-Frottier et al., 1995). The amplitude of the observed folds is well predicted by numerically determined scaling laws, which also predict amplitudes (400–500 km) consistent with tomographic images of slabs beneath some subduction zones (Ribe et al., 2007). On the other hand, when the slab is not too viscous, RTIs have time to develop from the slab resting on the interface and to sink into the lower mantle (Figure 70; Griffiths et al., 1995). When, however, the slab material reaches the depth of neutral buoyancy, it will spread at this level as a gravity current (Kerr and Lister, 1987, 1988). Slab deformation modes therefore finally depend on three parameters: the slab to mantle viscosity ratio $\gamma = \eta_S/\eta_L$, the ratio of sinking (Stokes) velocities in the upper and lower mantles ($\sim B_S\eta_U/\eta_L$), and the ratio of the slab's horizontal velocity to its vertical velocity. As the latter can change through time in nature, a slab probably passes through a number of different deformation modes during its history.

7.03.11.4.2 Free rollback velocity

If rollback controls in part the slab deformation at depth, what happens if the trench is free to move in response to the mantle flow? A number of systematic studies have recently been devoted to the dynamics of a slab sinking under its own weight (Figure 71). Regardless of the boundary conditions, subduction is always strongly time-dependent. After initiation, subduction accelerates as the mass of slab in the mantle increases (slab pull), while the slab dip is close to 90°. Then, subduction slows down after interaction with the 660 km interface, reaching a steady state when the slab spreads on the interface, followed by eventual penetration, accompanied by folding (Becker et al., 1999; Funiciello et al., 2003a,b, 2004, 2006, 2008; Schellart, 2004a,b, 2005, 2008, 2010). Besides, the amount of trench and slab rollback depends on the degree of lateral confinement of the flow (Funiciello et al., 2003a,b, 2004; Schellart, 2004a,b, 2005) and is much reduced in 2-D or when the sidewalls are within 600 km of the slab edges (Funiciello et al., 2003a,b). As in the experiments with rigid slabs (Buttles and Olson, 1998; Kincaid and Griffiths, 2003, 2004), slab rollback induces flow around the slab edges. The

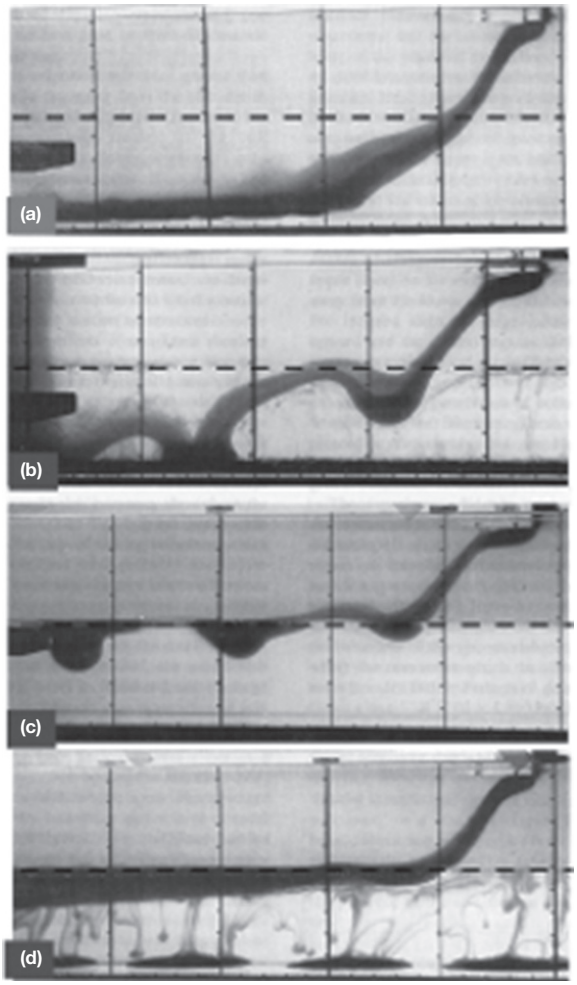


Figure 70 Different regimes of slab interaction with a viscosity/density interface for a slab, which is between 3 and 50 times more viscous than the lower layer. The trench velocity increases from (a) to (c) and the viscosity contrast from (c) to (d). Reproduced from Griffiths RW, Hackney RI, and van der Hilst RD (1995) A laboratory investigation of effects of trench migration on the descent of subducted slabs. *Earth and Planetary Science Letters* 133(1 2): 1–17.

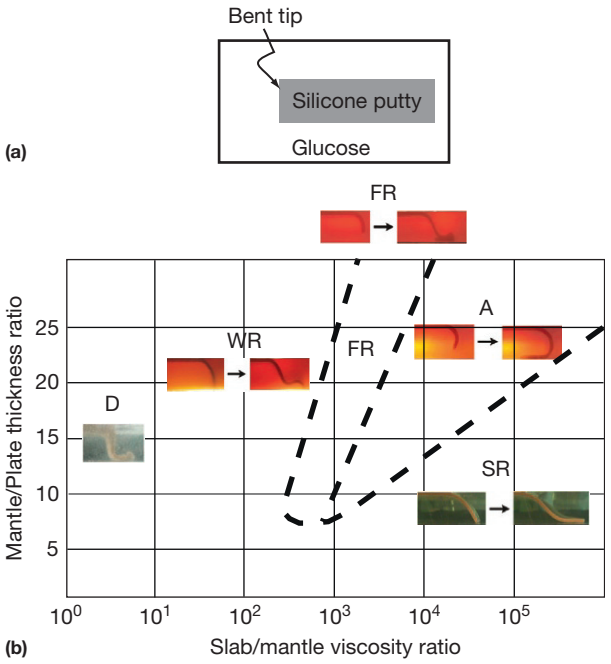


Figure 71 (a) Experimental setup used by the Roma III group. A slab of silicone putty is spread horizontally on glucose syrup. Subduction is initiated by forcing its tip 3 cm into the glucose. (b) Different subduction regimes encountered as a function of the slab/mantle viscosity ratios and thickness ratios. The photographs are courtesy of F. Funiciello. The thick dashed lines are the regime boundaries proposed by Schellart (2008).

lateral flow in turn forces the hinge line to adopt a convex shape toward the direction of retreat (Funiciello et al., 2003a,b; Schellart, 2004a,b; **Figure 68**). A large body of laboratory experiments (Bellahsen et al., 2005; Funiciello et al., 2006, 2008; Schellart, 2008, 2010) and numerical simulations (Li and Ribe, 2012; Stegman et al., 2010; Ribe, 2010) now exists, with variable thickness, width, viscosity, and density of the slab and mantle. The trench always retreats at the beginning of subduction, regardless of γ. However, the situation changes when the slab begins interacting with the lower boundary. The results are summarized in the regime diagram of **Figure 71(b)**. For very low viscosity ratios $\gamma < 100$, subduction occurs in a dripping (D) mode in which the sinking slab undergoes substantial deformation parallel to its midsurface (shortening/extension). For intermediate ratios $100 \leq \gamma \leq 10^4$, 3–5 different modes are seen depending on the slab thickness and width. In the trench retreating modes, the slab lies down more or less flat on the bottom boundary, and trench motion is retrograde, with a velocity smaller at lower γ (WR mode) than at higher γ (SR mode). In the folding (FR) mode, the slab folds once against the bottom boundary before continuing in a trench retreating mode. In the trench advancing (A) mode, the sinking slab bends over backward against the bottom boundary, and the trench advances. Some experiments can undergo an advancing episode when the slab tip encounters the bottom boundary, before entering the FR mode (Li and Ribe, 2012; Schellart, 2008). The boundaries between the different regimes proposed by Schellart (2008) on **Figure 71(b)** depend on slab width (Ribe, 2010; Schellart, 2008) and on preexisting mantle flow (Boutelier and Cruden, 2008). There is a very good agreement between laboratory experiments and numerical simulations (Li and Ribe, 2012; Ribe, 2010; Stegman et al., 2010). The same modes are also found when the plate velocity V_P is fixed at the ridge ('ridge push'; Funiciello et al., 2004; Schellart, 2005): a relatively low V_P results in relatively fast hinge retreat V_T with backward sinking of the slab and a backward draping slab geometry (WR). With increasing V_P, hinge migration is relatively small, resulting in subvertical sinking of the slab and a folded slab piling geometry (FR). For very high V_P, the hinge migrates forward, resulting in a forward draping slab geometry (A). These modes are characterized by different partitioning of the subduction into plate and trench motion. The lithospheric radius of curvature, which depends upon plate characteristics (stiffness and thickness) and the mantle thickness, seem to exert a primary control on the trench behavior (Bellahsen et al., 2005; Schellart, 2008). The subduction velocity results from the balance between acting and resisting forces. The bending radius of the slab increases linearly of the plate

thickness (Irvine and Schellart, 2012). The part of lithospheric bending in the energy dissipation is still debated. Some studies estimated that it represents 75–95% of the total resisting forces (Becker et al., 1999; Bellahsen et al., 2005) when others suggested a much lower contribution of 2–40% (Buffett, 2006; Leng and Zhong, 2010; Schellart, 2010). Recent experiments show that the energy dissipation ratio varies much during subduction, from a peak value ~70–80% shortly after the onset of subduction to a range of ~4–10% during steady-state subduction (Irvine and Schellart, 2012). Using FT image analysis, Funiciello et al. (2006) investigated the influence of plate width and mantle viscosity/density on rollback and the induced poloidal and toroidal components of the mantle circulation. The poloidal component is the response to the viscous coupling between the slab motion and the mantle, while the toroidal component is produced by lateral slab migration. The experiments show that both components are important from the beginning of the subduction process and strongly intermittent. Schellart (2010) looked at the influence of γ and the slab velocity on the trench curvature. The latter depended strongly on the subduction mode and on the toroidal motions located at the edges of the slab. The subduction regime can also be modified by the presence on the subducting lithosphere of ridges and plateaus (Martinod et al., 2005) or of continental crust (Bialas et al., 2011). On the other hand, the presence of an overriding plate does not seem to modify significatively the subduction mode. On the contrary, slab rollback is quite efficient at driving extension of the overriding plate (Meyer and Schellart, 2013).

The experiments have therefore established that mantle subduction is a strongly time-dependent and 3-D phenomenon. Hence, mantle flow in subduction zones cannot be correctly described by models assuming a 2-D steady-state process. Subduction dynamics involve not only slab pull and ridge push but also trench rollback and interaction with the 660 km depth discontinuity. All these effects explain the diversity of slab deformation and subduction history observed in the laboratory and in nature (e.g., Faccenna et al., 1999; Funiciello et al., 2008; Griffiths et al., 1995; Guillou-Frottier et al., 1995; Kincaid and Olson, 1987; Schellart, 2004a,b; Schellart et al., 2007; Stegman et al., 2010; Wu et al., 2008). Despite the multiplicity of regimes, time-dependent trench rollback is the most common occurrence, which explains well the formation of back-arc basins (e.g., Faccenna et al., 1999). The comparison of regional models with subduction laboratory and numerical experiments reveals that the range of observed upper mantle slab shapes, the existence of both retreating slabs and advancing slabs, the trench curvature, and the back-arc deformations require slabs to be only 150–500 times more viscous than the surrounding mantle (Funiciello et al., 2008; Ribe, 2010; Schellart, 2008; Schellart et al., 2007; Stegman et al., 2010). The challenge now is to obtain scaling laws to predict the complete regime diagram of subduction and the characteristics of each regime. This will require including thermal effects and addressing the problem of lithospheric rheology. All the laboratory experiments that have been done so far used a slab with Newtonian rheology. Another problem is to estimate the influence of surface tension, which is unavoidable in laboratory experiments with free surfaces (Jacoby, 1976) and which must be accounted for to obtain quantitative measurements of the forces resisting subduction and laboratory-based scaling laws for trench motion that can be applied to the mantle.

7.03.12 Conclusions

Quantitative analogue laboratory experiments have played an indispensable role in advancing the field of geodynamics since the earliest days of its existence as a discipline. Reflecting on this history helps one to understand more clearly what an experimental approach has to offer. At the top of any list must surely be the discovery of new phenomena. Laboratory experiments have always been the source of many of the most exciting new discoveries in fluid mechanics generally, and geodynamics itself offers numerous examples: a partial list would include the convective planforms in a variable-viscosity fluid, the diversity of thermal and compositional plumes, the entrainment in plumes, and the different modes of subduction. A second important contribution of the experimental approach is to the formation of influential new concepts and models. An obvious example here is the classical 'head plus tail' picture of a mantle plume; it is no exaggeration to say that this model had its origin almost entirely in the remarkable laboratory experiments of Whitehead and Luther (1975), Olson and Singer (1985), and Griffiths (1986a), among others, and in the beautiful photographic images that came from them. A third important role that laboratory experiments can play is to test theories arrived at using other methods: one thinks, for example, of the experiments performed by Whitehead and Luther (1975) to test the linear theory of the RTI. Experimental tests are especially important whenever the theory in question involves assumptions of an asymptotic nature or whose validity is not obvious.

The rapid evolution of high-performance numerical computing in recent years has provided still another role for experimental geodynamics: that of verifying and benchmarking complex numerical codes for phenomena such as convection with temperature-dependent viscosity (Busse et al., 1994), subduction (Schmeling et al., 2008; van Keken, 2008), plumes (van Keken, 1997; Vatteville et al., 2009), and convection with a chemical field (van Keken et al., 1997). Such benchmarks are particularly important for 3-D time-dependent flows with large variations in material properties, which are still difficult to treat reliably by numerical means. By the same token, the ability of numerical models to deliver detailed representations of 2-D and 3-D field variables has encouraged the development of powerful new techniques for quantitative flow visualization in the laboratory, such as the observation of the temperature field using TLCs (e.g., Davaille et al., 2011; Rhee et al., 1984), or the determination of the velocity and stretching fields by PIV (e.g., Adrian, 1991; Funiciello et al., 2006; Voth et al., 2002). This has led to an increasing number of studies using a combination of laboratory experiments and numerical simulations (e.g., Funiciello et al., 2008; Lohmann et al., 2009; Massmeyer et al., 2013; Stegman et al., 2010; van Keken et al., 2013; Vatteville et al., 2009).

While the future of experimental geodynamics is impossible to predict, it is not hard to identify some critical scientific questions where an experimental approach is likely to be

fruitful. One is the nature of entrainment and mixing between different mantle reservoirs, which involves phenomena occurring on short length scales and with little diffusion that are hard to resolve numerically. A second is the generation of plate tectonics by mantle convection, which cannot be understood in terms of traditional Newtonian fluid mechanics. Modeling this process in the laboratory requires new developments in the characterization and experimental deployment of fluids with complex rheology, and recent experiments using gels or colloids seem promising. But the study of convection in these 'soft' fluids is still in its infancy. It will be a necessary step though to decipher the first ages of the Earth and the passage from magma ocean to solid-state convection.

Acknowledgments

This work benefited from discussions with Judith Vatteville and from the comments of Dave Bercovici.

References

Abram C, Fond B, Heyes AL, and Beyrau F (2013) High-speed planar thermometry and velocimetry using thermographic phosphor particles. *Applied Physics B* 111: 155–160.

Acocella V (2008) Transform faults or Overlapping Spreading Centers? Oceanic ridge interactions revealed by analogue models. *Earth and Planetary Science Letters* 265: 379–385.

Adrian RJ (1983) Laser velocimetry. In: Goldstein RJ (ed.) *Fluid Mechanics Measurements*, pp. 155–244. Washington: Hemisphere.

Adrian RJ (1991) Particle imaging techniques for experimental fluid mechanics. *Annual Review of Fluid Mechanics* 23: 261–304.

Ahlers G (2001) Effect of sidewall conductance on heat-transport measurements for turbulent Rayleigh–Bénard convection. *Physical Review E* 62: 015303.

Ahlers G (2005) Experiments with Rayleigh-Bénard convection. In: Mu-tabazi I, Wesfreid Jose E, and Guyon E (eds.) *Dynamics of Spatio-Temporal Cellular Structures – Henri Bénard Centenary Review*. *Springer Tracts in Modern Physics*, vol. 207. New York: Springer.

Ahlers G, Grossmann S, and Lohse D (2009) Heat transfer and large scale dynamics in turbulent Rayleigh–Bénard convection. *Reviews of Modern Physics* 81: 503.

Albers M and Christensen UR (2001) Channeling of plume flow beneath mid-ocean ridges. *Earth and Planetary Science Letters* 187: 207–220.

Allegre CJ (1987) Isotope geodynamics. *Earth and Planetary Science Letters* 86: 175–203.

Androvandi S (2009) *Convection multi-chelles dans un Manteau dont la viscosit dpend de la temperature*. Ph.D. Thesis, Institut de Physique du Globe de Paris, France.

Androvandi S, Davaille A, Limare A, Fouquier A, and Marais C (2011) At least three scales of convection in a mantle with strongly temperature-dependent viscosity. *Physics of the Earth and Planetary Interiors* 188: 132–141. http://dx.doi.org/10.1016/j.pepi.2011.07.004.

Aref H (1984) Stirring by chaotic advection. *Journal of Fluid Mechanics* 143: 1–15.

Arndt N and Davaille A (2013) Episodic Earth evolution. *Tectonophysics* 609: 661–674.

Asaeda T and Watanabe K (1989) The mechanism of heat transport in thermal convection at high Rayleigh numbers. *Physics of Fluids A* 1: 861–867.

Asseban A, Lallemand M, Saulnier J-B, et al. (2000) Digital speckle photography and speckle tomography in heat transfer studies. *Optics and Laser Technology* 32: 583–592.

Assenheimer M and Steinberg V (1993) Rayleigh–Bénard convection near the gas–liquid critical point. *Physical Review Letters* 70: 3888–3891.

Bajolet F, Galeano J, Funiciello F, Moroni M, Negredo A-M, and Faccenna C (2012) Continental delamination: Insights from laboratory models. *Geochemistry, Geophysics, Geosystems* 13: Q02009. http://dx.doi.org/10.1029/2011GC003896.

Baker DJ (1966) A technique for the precise measurement of small fluid velocities. *Journal of Fluid Mechanics* 26: 573–575.

Balmforth NJ and Rust AC (2009) Weakly nonlinear viscoplastic convection. *Journal of Non-Newtonian Fluid Mechanics* 158: 36–45.

Batchelor GK (1954) Heat convection and buoyancy effects in fluids. *Quarterly Journal of the Royal Meteorological Society* 80: 339–358.

Baudry N and Kroenke L (1991) Intermediate-wavelength (400–600 km), South Pacific geoidal undulations: Their relationship to linear volcanic chains. *Earth and Planetary Science Letters* 102: 430–443.

Becker T, Faccenna C, Giardini D, and OConnell R (1999) The development of slabs in the upper mantle: Insights from numerical and laboratory experiments. *Journal of Geophysical Research* 104: 15207–15226.

Becker TW and Faccenna C (2009) A review of the role of subduction dynamics for regional and global plate motions. In: Lallemand S and Funiciello F (eds.) *Subduction Zone Geodynamics*, pp. 3–34. Springer.

Bellahsen N, Faccenna C, and Funiciello F (2005) Dynamics of subduction and plate motion in laboratory experiments: Insights into the plate tectonics behavior of the Earth. *Journal of Geophysical Research* 110: B01401. http://dx.doi.org/10.1029/2004JB002999.

Bénard H (1900) Les tourbillons cellulaires dans une nappe liquide. *Revue Generale Sciences de Pures et Appliquees* 11: 1261–1271, 1309–1328.

Bénard H (1901) Les tourbillons cellulaires dans une nappe liquide transportant la chaleur par convection en régime permanent. *Annales de Chimie Physique* 23: 62–144.

Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121.

Bercovici D and Mahoney J (1994) Double flood basalts and plume head separation at the 660-kilometer discontinuity. *Science* 266: 1367–1369.

Bercovici D and Kelly A (1997) Nonlinear initiation of diapirs and plume heads. *Physics of the Earth and Planetary Interiors* 101: 119–130.

Bercovici D, Ricard Y, and Richards MA (2000) The relation between mantle dynamics and plate tectonics: A primer. In: Richards MA, Gordon R, and Van der Hilst R (eds.) *The History and Dynamics of Global Plate Motions*. *AGU Geophysical Monograph*, vol. 21, pp. 5–46.

Bercovici D, Ricard Y, and Schubert G (2001a) A two-phase model for compaction and damage, 1. General theory. *Journal of Geophysical Research* 106: 8887–8906.

Bercovici D, Ricard Y, and Schubert G (2001b) A two-phase model for compaction and damage, 3. Applications to shear localization and plate boundary formation. *Journal of Geophysical Research* 106: 8925–8940.

Bercovici D and Ricard Y (2012) Mechanisms for the generation of plate tectonics by two-phase grain-damage and pinning. *Physics of the Earth and Planetary Interiors* 202–203: 27–55.

Bercovici D and Ricard Y (2013) Plate generation with two-phase grain-damage and pinning: Source-sink model and toroidal Flow. *Earth and Planetary Science Letters* 365: 275–288. http://dx.doi.org/10.1016/j.epsl.2013.02.002.

Berner H, Ramberg H, and Stephansson O (1972) Diapirism in theory and experiment. *Tectonophysics* 15: 197–218.

Bialas RW, Funiciello F, and Faccenna C (2011) Subduction and exhumation of continental crust: Insights from laboratory models. *Geophysical Journal International* 184: 43–64.

Billen MI, Gurnis M, and Simons M (2003) Multiscale dynamics of the Tonga Kermadec subduction zone. *Geophysical Journal International* 153: 359–388.

Billen MI (2008) Modeling the dynamics of subducting slabs. *Annual Reviews of Earth and Planetary* 36: 325–356.

Biot MA and Ode H (1965) Theory of gravity instability with variable overburden and compaction. *Geophysics* 30: 213–227.

Bodenschatz E, Gemelke N, Carr J, and Ragnarsson R (1997) Rifts in spreading wax layers: An analogy to the mid-ocean rift formation. *Symposium on Localization Phenomena and Dynamics of Brittle and Granular System*. Columbia University.

Bodenschatz E, Pesch W, and Ahlers G (2000) Recent developments in Rayleigh–Bénard convection. *Annual Review of Fluid Mechanics* 32: 709–773.

Bonatti E (1985) Punctiform initiation of seafloor spreading in the Red Sea during transition from a continental to an oceanic rift. *Nature* 316: 33–37.

Booker JR (1976) Thermal convection with strongly temperature dependent viscosity. *Journal of Fluid Mechanics* 76: 741–754.

Booker JR and Stengel KC (1978) Further thoughts on convective heat transport in a variable viscosity fluid. *Journal of Fluid Mechanics* 86: 289–291.

Boss AP and Sacks IS (1986) High spatial resolution models of time-dependent, layered mantle convection. *Geophysical Journal of the Royal Astronomical Society* 87: 241–264.

Boutelier DA and Cruden AR (2008) Impact of regional mantle flow on subducting plate geometry and interplate stress: Insights from physical modelling. *Geophysical Journal International* 174: 719–732.

Boxman RL and Shlien DJ (1978) Interferometric measurement technique for the temperature field of axisymmetric buoyant phenomena. *Applied Optics* 17: 2788–2793.

Boxman RL and Shlien DJ (1981) Laminar starting plume temperature field measurement. *Journal of Heat and Mass Transfer* 24: 919–931.

Brown E, Nikilaenko A, Funfschilling D, and Ahlers G (2005) Heat transport in turbulent Rayleigh–Bénard convection: Effect of finite top- and bottom-plate conductivities. *Physics of Fluids* 17: 075108.

Buck WR (1987) Analysis of the cooling of a variable viscosity fluid with applications to the Earth. *Geophysical Journal of the Royal Astronomical Society* 89: 549–577.

Buffett BA (2006) Plate force due to bending at subduction zones. *Journal of Geophysical Research* 111: B09405 http://dx.doi.org/10.1029/2006JB004295.

Busse FH (1967a) On the stability of two-dimensional convection in a layer heated from below. *Journal of Mathematical Physics* 46: 140–149.

Busse FH (1967b) The stability of finite amplitude cellular convection and its relation to an extremum principle. *Journal of Fluid Mechanics* 30: 625–635.

Busse FH and Whitehead JA (1971) Instabilities of convection rolls in high Prandtl number fluid. *Journal of Fluid Mechanics* 47: 305–320.

Busse FH (1978) Non linear properties of thermal convection. *Reports on Progress in Physics* 41: 1929–1967.

Busse FH (1981) On the aspect ratios of two-layer mantle convection. *Physics of the Earth and Planetary Interiors* 24: 320–324.

Busse FH (1989) Fundamentals of thermal convection. In: Perltier WR (ed.) *Mantle Convection: Plate Tectonics and Global Dynamics*, pp. 23–95. Montreux: Gordon and Breach.

Busse FH, Christensen U, Clever R, et al. (1994) 3D convection at infinite Prandtl number in cartesian geometry – A benchmark comparison. *Geophysical and Astrophysical Fluid Dynamics* 75: 39–59.

Busse FH (2012) Viewpoint: The twins of turbulence research. *Physics* 5: 4–12.

Buttles J and Olson P (1998) A laboratory model of subduction zone anisotropy. *Earth and Planetary Science Letters* 164: 245–262.

Cadio C, Panet I, Davaille A, Diament M, Mtivier L, and de Viron O (2011) Pacific geoid anomalies revisited in light of thermochemical oscillating domes in the lower mantle. *Earth and Planetary Science Letters* 306: 123–135.

Campbell IH and Griffiths RW (1990) Implications of mantle plume structure for the evolution of flood basalts. *Earth and Planetary Science Letters* 99: 79–93.

Canright D and Morris S (1993) Buoyant instability of a viscous film over a passive fluid. *Journal of Fluid Mechanics* 255: 349–372.

Cardin P and Nataf H-C (1991) Nonlinear dynamical coupling observed near the threshold of convection in a two-layer system. *Europhysics Letters* 14: 655–660.

Carrigan CR (1982) Multiple scale convection in the Earth's mantle: A three dimensional study. *Science* 215: 965–967.

Castaing B, Gunaratne G, Heslot F, et al. (1989) Scaling of hard turbulence in Rayleigh–Bénard convection. *Journal of Fluid Mechanics* 204: 1–30.

Cazenave A, Monnerau M, and Gibert D (1987) Seasat gravity undulations in the central Indian ocean. *Physics of the Earth and Planetary Interiors* 48: 130–141.

Chandrasekhar S (1977) *Liquid Crystals*, p. 342. Cambridge: Cambridge University Press.

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. New York: Dover.

ChavanneX Chill F, Chabaud F, Castaing B, and Hebral B (2001) Turbulent Rayleigh–Bénard convection in gaseous and liquid He. *Physics of Fluids* 13: 1300–1320.

Chay A and Shlien DJ (1986) Scalar field measurements of a laminar starting plume cap using digital processing of interferograms. *Physics of Fluids* 29: 2358–2366.

Chemenda AI, Burg JP, and Mattauer M (2000) Evolutionary model of the Himalaya–Tibet system: Geopoem based on new modelling, geological and geophysical data. *Earth and Planetary Science Letters* 174: 397–409.

Chen F and Chen CF (1989) Experimental investigation of convective stability in a superposed fluid and porous layer when heated from below. *Journal of Fluid Mechanics* 207: 311–321.

Chen MM and Whitehead JA (1968) Evolution of two-dimensional periodic Rayleigh convection cells of arbitrary wave-numbers. *Journal of Fluid Mechanics* 31: 1–15.

Chilla F, Rastello M, Chaumat S, and Castaing B (2004) Long relaxation times and tilt sensitivity in Rayleigh–Bénard turbulence. *European Physical Journal B* 40: 223–227.

Chilla F and Schumacher J (2012) New perspectives in turbulent Rayleigh–Bénard convection. *The European Physical Journal E* 35: 58–83.

Choblet G and Sotin C (2000) 3D thermal convection with variable viscosity: Can transient cooling be described by a quasi-static scaling law? *Physics of the Earth and Planetary Interiors* 119: 321–336.

Christensen U (1984) Instability in a hot boundary layer and initiation of thermochemical plumes. *Annals of Geophysics* 2: 311–320.

Christensen U and Yuen DA (1984) The interaction of a subducting slab with a chemical or phase boundary. *Journal of Geophysical Research* 89: 4389–4402.

Christensen U (1985) Heat transport by variable viscosity convection II: Pressure influence, non-Newtonian rheology and decaying heat sources. *Physics of the Earth and Planetary Interiors* 37: 183–205.

Christensen UR and Harder H (1991) 3-D Convection with variable viscosity. *Geophysical Journal International* 104: 213–226.

Christensen UR and Hofmann AW (1994) Segregation of subducted oceanic crust in the convecting mantle. *Journal of Geophysical Research* 99: 19867–19884.

Ciofalo M, Signorino M, and Simiano M (2003) Tomographic particle-image velocimetry and thermography in Rayleigh–Bénard convection using suspended thermochromic liquid crystals and digital image processing. *Experiments in Fluids* 34: 156–172.

Condie KC (1998) Episodic continental growth and supercontinents: A mantle avalanche connection? *Earth and Planetary Science Letters* 163: 97–108.

Condie KC, Belousova E, Griffin WL, and Sircombe KN (2009) Granitoid events in space and time: Constraints from igneous and detrital zircon age spectra. *Gondwana Research* 15: 228–242.

Cottrell E, Jaupart C, and Molnar P (2004) Marginal stability of thick continental lithosphere. *Geophysical Research Letters* 31. http://dx.doi.org/10.1029/2004GL020332.

Coulliette DL and Loper DE (1995) Experimental, numerical and analytical models of mantle starting plumes. *Physics of the Earth and Planetary Interiors* 92: 143–167.

Count Rumford (1798) Heat is a form of motion: An experiment in boring cannon. *Philosophical Transactions* 88.

Crane K (1985) The spacing of rift axis highs: Dependence upon diapiric processes in the underlying asthenosphere? *Earth and Planetary Science Letters* 72: 405–414.

Cruickshank JO and Munson BR (1981) Viscous fluid buckling of plane and axisymmetric jets. *Journal of Fluid Mechanics* 113: 221–239.

Cserepes L and Rabinowicz M (1985) Gravity and convection in a two-layered mantle. *Earth and Planetary Science Letters* 76: 193–207.

Cserepes L, Rabinowicz M, and Rosemberg-Borot C (1988) Three-dimensional infinite Prandtl number convection in one and two layers with implications for the Earth's gravity field. *Journal of Geophysical Research* 93: 12009–12025.

Dabiri D and Gharib M (1991) Digital particle image thermometry: The method and implementation. *Experiments in Fluids* 11: 77–86.

Dabiri D and Gharib M (1996) The effects of forced boundary conditions on flow within a cubic cavity using digital particle image thermometry and velocimetry (DPITV). *Experimental Thermal and Fluid Science* 13: 349–363.

Dalziel SB, Hughes GO, and Sutherland BR (2000) Whole field density measurements by "synthetic schlieren" *Experiments in Fluids* 28: 322–345.

Darbouli M, Metivier C, Piau J-M, Magnin A, and Abdelali A (2013) Rayleigh–Bénard convection for viscoplastic fluids. *Physics of Fluids* 25: 023101.

Davaille A and Jaupart C (1993) Transient high-Rayleigh number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.

Davaille A and Jaupart C (1994) Onset of thermal convection in fluids with temperature-dependent viscosity: Application to the oceanic mantle. *Journal of Geophysical Research* 99: 19853–19866.

Davaille A (1999a) Two-layer thermal convection in viscous fluids. *Journal of Fluid Mechanics* 379: 223–253.

Davaille A (1999b) Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* 402: 756–760.

Davaille A, Girard F, and Le Bars M (2002) How to anchor plumes in a convecting mantle? *Earth and Planetary Science Letters* 203: 621–634.

Davaille A, Le Bars M, and Carbonne C (2003) Thermal convection in a heterogeneous mantle. *Comptes rendus Geoscience* 335: 141–156.

Davaille A and Vatteville J (2005) On the transient nature of mantle plumes. *Geophysical Research Letters* 32. http://dx.doi.org/10.1029/2005GL023029.

Davaille A, Stutzmann E, Silveira G, Besse J, and Courtillot V (2005) Convective patterns under the Indo-Atlantic "box". *Earth and Planetary Science Letters* 239: 233–252.

Davaille A, Androvandi S, Vatteville J, Limare A, Vidal V, and Lebars M (2008) Thermal boundary layer instabilities in viscous fluids. In: *ISFV13 – 13th International symposium on flow visualization, July 1–4, 2008, Nice, France*, pp. 12–18.

Davaille A, Limare A, Touitou F, Kumagai I, and Vatteville J (2011) Anatomy of a laminar starting plume at high Prandtl number. *Experiments in Fluids* 50: 285–300.

Davaille A, Gueslin B, Massmeyer A, and Di Giuseppe E (2013) Thermal instabilities in yield stress fluids: Existence and morphology. *Journal of Non-Newtonian Fluid Mechanics* 193: 144–153.

Davies GF (1988) Ocean bathymetry and mantle convection, 1. Large-scale flow and hotspots. *Journal of Geophysical Research* 93: 10467–10480.

Davies GF (1995) Punctuated tectonic evolution of the Earth. *Earth and Planetary Science Letters* 136: 363–379.

Davy P and Cobbold PR (1991) Experiments on shortening of a 4-layer continental lithosphere. *Tectonophysics* 188: 1–25.
Deardorff JW, Willis GE, and Lilly DK (1969) Laboratory investigation of non-steady penetrative convection. *Journal of Fluid Mechanics* 35: 7–31.
Deschamps F, Kaminski E, and Tackley P (2011) A deep mantle origin for the primitive signature of ocean island basalt. *Nature Geoscience* 4: 879–882.
de Gennes P and Prost J (1993) *The Physics of Liquid Crystals*, 2nd edn., p. 597. Oxford: Clarendon Press.
Di Giuseppe E, Funiciello F, Corbi F, Ranalli G, and Mojoli G (2009) Gelatins as rock analogs: A systematic study of their rheological and physical properties. *Tectonophysics* 473: 391–403.
Di Giuseppe E and Davaille A (2010) Episodic Subduction arising from convection in a complex-rheology fluid: Insights from laboratory experiments. *Geophysical Research Abstracts* 12: 07954.
Di Giuseppe E, Davaille A, Mittelstaedt E, and Franois M (2012) Rheological and mechanical properties of silica colloids: From liquid Newtonian to brittle behavior. *Rheologica Acta* 51: 451–465.
Dubuffet F, Rabinowicz M, and Monnereau M (2000) Multiple scales in mantle convection. *Earth and Planetary Science Letters* 178: 351–366.
Dumoulin C, Doin M-P, and Fleitout L (2001) Numerical simulations of the cooling of an oceanic lithosphere above a convecting mantle. *Physics of the Earth and Planetary Interiors* 125: 45–64.
Druken KA, Long MD, and Kincaid C (2011) Patterns in seismic anisotropy driven by rollback subduction beneath the High Lava Plains. *Geophysical Research Letters* 38: L13310. http://dx.doi.org/10.1029/2011GL047541.
Eberle MA and Forsyth DW (1995) Regional viscosity variations, small-scale convection and slope of the depth-age1/2 curve. *Geophysical Research Letters* 22: 473–476.
Egbers Ch, Beyer W, Bonhage A, Hollerbach R, and Beltrame P (2003) The GEOFLOW-experiment on ISS (part I): Experimental preparation and design. *Advances in Space Research* 32: 171–180.
Ekstrm G and Dziewonski AM (1998) The unique anisotropy of the pacific upper mantle. *Nature* 394: 168–172.
Elder JW (1967) Convective self-propulsion of continents. *Nature* 214: 657–660.
Elder JW (1968) The unstable thermal interface. *Journal of Fluid Mechanics* 32: 69–96.
Espurt N, Funiciello F, Martinod J, et al. (2008) Flat subduction dynamics and deformation of the South American plate: Insights from analog modeling. *Tectonics* 27: TC3011. http://dx.doi.org/10.1029/2007TC002175.
Faccenna C, Davy P, Brun J-P, et al. (1996) The dynamics of back-arc extension: An experimental approach to the opening of the Tyrrhenian Sea. *Geophysical Journal International* 126: 781–795.
Faccenna C, Giardini D, Davy P, and Argentieri A (1999) Initiation of subduction at Atlantic-type margins: Insights from laboratory experiments. *Journal of Geophysical Research* 104: 2749–2766.
Faccenna C, Piromallo C, Crespo-Blanc A, Jolivet L, and Rossetti F (2004) Lateral slab deformation and the origin of the western Mediterranean arcs. *Tectonics* 23: TC1012. http://dx.doi.org/10.1029/2002TC001488.
Farnetani CG and Richards MA (1995) Thermal entrainment and melting in mantle plumes. *Earth and Planetary Science Letters* 136: 251–267.
Farnetani CG (1997) Excess temperature of mantle plumes: The role of chemical stratification across D″. *Geophysical Research Letters* 24: 1583–1586.
Fedotov SA (1975) Mechanism of magma ascent and deep feeding channels of island arc volcanoes. *Bulletin of Volcanology* 39: 241–254.
Feighner MA and Richards MA (1995) The fluid dynamics of plume–ridge and plume-plate interactions: An experimental investigation. *Earth and Planetary Science Letters* 129: 171–182.
Feighner MA, Kellogg LH, and Travis BJ (1995) Numerical modeling of chemically buoyant plumes at spreading ridges. *Geophysical Research Letters* 22: 715–718.
Fialko Y and Pearse J (2012) Sombrero uplift above the Altiplano-Puna magma body: Evidence of a ballooning mid-crustal diapir. *Science* 338: 250–252.
Fleitout L and Yuen DA (1984) Steady-state, secondary convection beneath lithospheric plates with temperature and pressure-dependent viscosity. *Journal of Geophysical Research* 89: 9227–9244.
Fletcher RC (1972) Application of mathematical model to emplacement of mantle gneiss domes. *American Journal of Science* 272: 197–216.
Fountain GO, Khakhar DV, and Ottino JM (1998) Visualization of three dimensional chaos. *Nature* 281: 683–686.
Fountain GO, Khakhar DV, Mezic I, and Ottino JM (2000) Chaotic mixing in a bounded three-dimensional flow. *Journal of Fluid Mechanics* 417: 265–301.
Fourel L, Milelli L, Jaupart C, and Limare A (2013) Generation of continental rifts, basins, and swells by lithosphere instabilities. *Journal of Geophysical Research - Solid Earth* 118: 2169–9356.
Fuji T (1963) Theory of the steady laminar natural convection above a horizontal line heat source and a point heat source. *International Journal of Heat and Mass Transfer* 6: 597–606.
Fujisawa N and Adrian R (1999) Three-dimensional temperature measurement in turbulent thermal convection by extended range scanning liquid crystal thermometry. *Journal of Visualization* 1: 355–364.
Fujisawa N and Funatani S (2000) Simultaneous measurement of temperature and velocity in a turbulent thermal convection by the extended range liquid crystal visualization technique. *Experiments in Fluids* 29: 158–165.
Fujisawa N and Hashizume Y (2001) An uncertainty analysis of temperature and velocity measured by a liquid crystal visualization technique. *Measurement Science and Technology* 12: 1235–1242.
Funiciello F, Morra G, Giardini D, and Regenauer-Lieb K (2003a) Dynamics of retreating slabs, part 1: Insights from 2-D numerical experiments. *Journal of Geophysical Research* 108: 2206. http://dx.doi.org/10.1029/2001JB000898.
Funiciello F, Faccenna C, Giardini D, and Regenauer-Lieb K (2003b) Dynamics of retreating slabs: 2. Insights from three-dimensional laboratory experiments. *Journal of Geophysical Research* 108: 2207.
Funiciello F, Faccenna C, and Giardini D (2004) Role of lateral mantle flow in the evolution of subduction system: Insights from 3-D laboratory experiments. *Geophysical Journal International* 157: 1393–1406.
Funiciello F, Moroni M, Piromallo C, Faccenna C, Cenedese A, and Bui HA (2006) Mapping mantle flow during retreating subduction: Laboratory models analyzed by feature tracking. *Journal of Geophysical Research* 111: B03402. http://dx.doi.org/10.1029/2005JB003792.
Funiciello F, Faccenna C, Heuret A, Di Giuseppe E, Lallemand S, and Becker TW (2008) Trench migration, net rotation and slab-mantle coupling. *Earth and Planetary Science Letters* 271: 233–240.
Funfschilling D, Brown E, Nikolaenko A, and Ahlers G (2005) Heat transport by turbulent Rayleigh–Bénard convection in cylindrical cells with aspect ratio one and larger. *Journal of Fluid Mechanics* 536: 145–155.
Futterer B, Egbers Ch, Koch S, Dahley N, and Jehring L (2010) First identification of sub- and supercritical convection patterns from 'GeoFlow', the geophysical flow simulation experiment integrated in Fluid Science Laboratory. *Acta Astronautica* 66: 193–200.
Futterer B, Dahley N, Koch S, Scurtu N, and Egbers Ch (2012) From isoviscous convective experiment GeoFlow I to temperature-dependent viscosity in GeoFlow II – Fluid physics experiments on-board ISS for the capture of convection phenomena in Earth's outer core and mantle. *Acta Astronautica* 71: 11–19.
Futterer B, Krebs A, Plesa A-C, et al. (2013) Sheet-like and plume-like thermal flow in a spherical convection experiment performed under microgravity. *Journal of Fluid Mechanics* 735: 647–683.
Gasteuil Y, Shew WL, Gibert M, Chilla F, Castaing B, and Pinton J-F (2007) Lagrangian temperature, velocity, and local heat flux measurement in Rayleigh-Bénard convection. *Physical Review Letters* 99: 234–302.
Gebhart B, Pera L, and Schorr AW (1970) Steady laminar natural convection plumes above a horizontal line heat source. *International Journal of Heat and Mass Transfer* 13: 161–171.
Geitling AV (1998) *Rayleigh–Bénard Convection*. Singapore: World Scientific.
Gerya T (2012) Origin and models of oceanic transform faults. *Tectonophysics* 522–523: 34–54.
Giannandrea E and Christensen U (1993) Variable viscosity convection experiments with a stress-free upper boundary and implications for the heat transport in the Earth's mantle. *Physics of the Earth and Planetary Interiors* 78: 139–152.
Globe S and Dropkin D (1959) Natural-convection heat transfer in liquids confined by two horizontal plates and heated from below. *Journal of Heat Transfer* 31: 24–28.
Gluckman B, Willaime H, and Gollub J (1993) Geometry of isothermal and isoconcentration surfaces in thermal turbulence. *Physics of Fluids A* 5: 647–651.
Göğüş OH, Pysklywec RN, Corbi F, and Faccenna C (2011) The surface tectonics of mantle lithosphere delamination following ocean lithosphere subduction: Insights from physical-scaled analogue experiments. *Geochemistry, Geophysics, Geosystems* 12: Q05004. http://dx.doi.org/10.1029/2010GC003430.
Goldstein RJ, Chiang HD, and See DL (1990) High Rayleigh-number convection in a horizontal enclosure. *Journal of Fluid Mechanics* 213: 111–126.
Gonnermann HM, Manga M, and Jellinek AM (2002) Dynamics and longevity of an initially stratified mantle. *Geophysical Research Letters* 29. http://dx.doi.org/10.1029/2002GL01485.
Gonnermann HM, Jellinek AM, Richards MA, and Manga M (2004) Modulation of mantle plumes and heat flow at the core mantle boundary by plate-scale flow: Results from laboratory experiments. *Earth and Planetary Science Letters* 226: 53–67.

Goren L, Aharonov E, Mulugeta G, Koyi HA, and Mart Y (2008) Ductile deformation of passive margins: A new mechanism for subduction initiation. *Journal of Geophysical Research* 113: B08411. http://dx.doi.org/10.1029/2005JB004179.

Grasset O and Parmentier EM (1998) Thermal convection in a volumetrically, infinite Prandtl number fluid with strongly temperature-dependent viscosity: Implications for planetary evolution. *Journal of Geophysical Research* 103: 18171–18181.

Griffiths RW (1986a) Thermals in extremely viscous fluids, including the effects of temperature-dependent viscosity. *Journal of Fluid Mechanics* 166: 115–138.

Griffiths RW (1986b) The differing effects of compositional and thermal buoyancy on the evolution of mantle diapirs. *Physics of the Earth and Planetary Interiors* 43: 261–273.

Griffiths RW and Turner JS (1988) Folding of viscous plumes impinging on a density or viscosity interface. *Geophysical Journal* 95: 397–419.

Griffiths RW and Richards MA (1989) The adjustment of mantle plumes to changes in plate motion. *Geophysical Research Letters* 16: 437–440.

Griffths RW, Gurnis M, and Eitelberg G (1989) Holographic measurements of surface topography in laboratory models of mantle hotspots. *Geophysical Journal* 96: 1–19.

Griffiths RW and Campbell IH (1990) Stirring and structure in mantle starting plumes. *Earth and Planetary Science Letters* 99: 66–78.

Griffiths RW and Campbell IH (1991) Interaction of mantle plume heads with the Earth's surface and the onset of small-scale convection. *Journal of Geophysical Research* 96: 18295–18310.

Griffiths RW, Hackney RI, and van der Hilst RD (1995) A laboratory investigation of effects of trench migration on the descent of subducted slabs. *Earth and Planetary Science Letters* 133(1–2): 1–17.

Grign C, Labrosse S, and Tackley PJ (2005) Convective heat transfer as a function of wavelength: Implications for the cooling of the Earth. *Journal of Geophysical Research* 110: B03409. http://dx.doi.org/10.1029/2004JB003376.

Grossmann S and Lohse D (2000) Scaling in thermal convection: A unifying theory. *Journal of Fluid Mechanics* 407: 27–56.

Grossmann S and Lohse D (2011) Multiple scaling in the ultimate regime of thermal convection. *Physics of Fluids* 23: 045108.

Guillaume B, Moroni M, Funiciello F, Martinod J, and Faccenna C (2010) Mantle flow and dynamic topography associated with slab window opening: Insights from laboratory models. *Tectonophysics* 496(1–4): 83–98.

Guillou L and Jaupart C (1995) On the effect of continents on mantle convection. *Journal of Geophysical Research* 100: 24217–24238.

Guillou-Frottier L, Buttles J, and Olson P (1995) Laboratory experiments on structure of subducted lithosphere. *Earth and Planetary Science Letters* 133: 19–34.

Gurnis M (1988) Large-scale mantle convection and the aggregation and dispersal of supercontinents. *Nature* 335: 695–699.

Gurnis M and Davies GF (1986) The effect of depth-dependent viscosity on convective mixing in the mantle and the possible survival of primitive mantle. *Geophysical Research Letters* 13: 541–544.

Gurnis M, Zhong S, and Toth J (2000) On the competing roles of fault reactivation and brittle failure in generating plate tectonics from mantle convection. In: Richards MA, Gordon R, and van der Hilst R (eds.) *The History and Dynamics of Global Plate Motions*, *AGU Geophysical Monograph 21*, pp. 73–94.

Haller G (2001) Finding finite-time invariant manifolds in two-dimensional velocity fields. *Chaos* 10: 99–108.

Haller G and Yuan G (2000) Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D* 147: 352–370.

Hanenkamp A and Merzkirch W (2005) Investigation of the properties of a sharp-focusing schlieren system by means of Fourier analysis. *Optics and Lasers in Engineering* 44: 159–169.

Hansen U and Yuen DA (1988) Numerical simulations of thermo-chemical instabilities at the core–mantle boundary. *Nature* 33: 237–240.

Hansen U and Yuen DA (2000) Extended-Boussinesq thermal-chemical convection with moving heat sources and variable viscosity. *Earth and Planetary Science Letters* 176: 401–411.

Haramina T and Tilgner A (2004) Coherent structures in boundary layers of Rayleigh–Bénard convection. *Physical Review E* 69: 056306.

Hartlep T, Tilgner A, and Busse FH (2003) Large scale structures in Rayleigh–Bénard convection at high Rayleigh numbers. *Physical Review Letters* 91: 064501.

Hauf W and Grigull U (1970) Optical methods in heat transfer. In: Hartnett JP and Irvine TF Jr. (eds.) *Advances in Heat Transfer*, pp. 133–366. New York: Academic.

Haxby WF and Weissel JK (1986) Evidence for small-scale convection from seasat altimeter data. *Journal of Geophysical Research* 91: 3507–3520.

Hawkesworth CJ, Cawood P, Kemp AIS, Storey CD, and Dhuime B (2009) A matter of preservation. *Science* 49–50.

Heflinger LO, Wuerker RF, and Brooks RE (1966) Holographic interferometry. *Journal of Applied Physics* 37: 642–649.

Herrick DL and Parmentier EM (1994) Episodic large-scale overturn of two-layer mantles in terrestrial planets. *Journal of Geophysical Research* 99: 2053–2062.

Heslot F, Castaing B, and Libchaber A (1987) Transitions to turbulence in helium gas. *Physical Review A* 36: 5870–5873.

Heutmaker MS and Gollub JP (1987) Wave-vector field of convective flow patterns. *Physical Review A* 35: 242–260.

Hieronymus CF and Bercovici D (1999) Discrete alternating hot-spot islands formed by interaction of magma transport and lithospheric flexure. *Nature* 397: 604–607.

Hofmann AW and Hart SR (1978) An assessment of local and regional isotopic equilibrium in the mantle. *Earth and Planetary Science Letters* 38: 44–62.

Holmes A (1931) Radioactivity and Earth movements. *The Geological Society of Glasgow* 18: 559–606.

Horsmann M and Merzkirch W (1981) Scattered light streaming birefringence in colloidal solutions. *Rheologica Acta* 20: 501–510.

Houseman GA (1990) The thermal structure of mantle plumes: Axisymmetric or triple-junction? *Geophysical Journal International* 102: 25–43.

Houseman GA and Molnar P (1997) Gravitational (Rayleigh–Taylor) instability of a layer with non-linear viscosity and convective thinning of continental lithosphere. *Geophysical Journal International* 128: 125–150.

Houseman GA and Gubbins D (1997) Deformation of subducted oceanic lithosphere. *Geophysical Journal International* 131: 535–551.

Howard LN (1964) Convection at high Rayleigh number. In: Grtler H (ed.) *Applied Mechanics*. *Proceedings of 11th International Congress*. Berlin: Springer.

Huang J, Zhong S, and van Hunen J (2003) Controls on sublithospheric small-scale convection. *Journal of Geophysical Research* 108(B8): 2405. http://dx.doi.org/10.1029/2003JB002456.

Irvine DN and Schellart WP (2012) Effect of plate thickness on bending radius and energy dissipation at the subduction zone hinge. *Journal of Geophysical Research* 117: B06405. http://dx.doi.org/10.1029/2011JB009113.

Ito G (2001) Reykjanes 'V'-shaped ridges originating from a pulsing and dehydrating mantle plume. *Nature* 411: 681–684.

Jacoby WR (1970) Instability in the upper mantle and global plate movements. *Journal of Geophysical Research* 75: 5671–5680.

Jacoby WR (1973) Model experiment of plate movements. *Nature Physical Sciences* 242: 130–134.

Jacoby WR (1976) Paraffin model experiment of plate tectonics. *Tectonophysics* 35: 103–113.

Jacoby WR and Schmeling H (1982) On the effects of the lithosphere on mantle convection and evolution. *Physics of the Earth and Planetary Interiors* 29: 305–319.

Jackson MPA and Talbot CJ (1989) Anatomy of mushroom shaped diapirs. *Journal of Structural Geology* 11: 211–230.

Jaupart C, Brandeis G, and Allègre CJ (1984) Stagnant layers at the bottom of convecting magma chambers. *Nature (London)* 308: 535–538.

Jaupart C and Brandeis G (1986) The stagnant bottom layer of convecting magma chambers. *Earth and Planetary Science Letters* 80: 183–199.

Jaupart C and Parsons B (1985) Convective instabilities in a variable viscosity fluid cooled from above. *Physics of the Earth and Planetary Interiors* 39: 32–41.

Jaupart C, Mareschal J-C, Guillou-Frottier L, and Davaille A (1998) Heat flow and thickness of the lithosphere in the Canadian Shield. *Journal of Geophysical Research* 103: 15269–15286.

Jaupart C, Molnar P, and Cottrell E (2007) Instability of a chemically dense layer heated from below and overlain by a deep less viscous fluid. *Journal of Fluid Mechanics* 572: 433–469.

Javoy M (1999) Chemical Earth models. *Comptes Rendus de l'Acadêmie des Sciences* 329: 537–555.

Jellineck AM, Kerr RC, and Griffiths RW (1999) Mixing and compositional stratification produced by natural convection. 1. Experiments and their application to Earth's core and mantle. *Journal of Geophysical Research* 104: 7183–7201.

Jellinek AM and Manga M (2002) The influence of a chemical boundary layer on the fixity, spacing and lifetime of mantle plumes. *Nature* 41: 760–763.

Jellinek AM, Lenardic A, and Manga M (2002) The influence of interior mantle temperature on the structure of plumes: Heads for Venus, Tails for the Earth. *Geophysical Research Letters* 29. http://dx.doi.org/10.1029/2001GL014624.

Jellinek AM, Gonnermann HM, and Richards MA (2003) Plume capture by divergent plate motions: Implications for the distribution of hotspots, geochemistry of mid-ocean ridge basalts, and heat flux at the core–mantle boundary. *Earth and Planetary Science Letters* 205: 367–378.

Jellinek AM and Manga M (2004) Links between long-lived hotspots, mantle plumes, D, and plate tectonics. *Reviews of Geophysics* 42. http://dx.doi.org/10.1029/2003RG000144.

Jellinek AM and Lenardic A (2009) Effects of spatially varying roof cooling on Rayleigh–Benard convection in a fluid with temperature-dependent viscosity. *Journal of Fluid Mechanics* 629: 109–137.
Jousselin D and Nicolas A (2000) Oceanic ridge off-axis deep structure in the mansah region (Sumail massif, Oman ophiolite). *Marine Geophysical Research* 21: 243–257.
Jurine D, Jaupart C, Brandeis G, and Tackley PJ (2005) Penetration of mantle plumes through depleted lithosphere. *Journal of Geophysical Research* 110. http://dx.doi.org/10.1029/2005JB003751.
Karlstrom L and Manga M (2006) Origins and implications of zigzag rift patterns on lava lakes. *Journal of Volcanology and Geothermal Research* 154: 317–324.
Kaminski E and Jaupart C (2003) Laminar starting plumes in high-Prandtl-number fluids. *Journal of Fluid Mechanics* 478: 287–298.
Katsaros KB, Liu WT, Businger JA, and Tillman JE (1977) Heat thermal structure in the interfacial boundary layer measured in an open tank of water in turbulent free convection. *Journal of Fluid Mechanics* 83: 311–335.
Katz RF, Ragnarsson R, and Bodenschatz E (2005) Tectonic microplates in a wax model of sea floor spreading. *New Journal of Physics* 7. http://dx.doi.org/10.1088/1367-2630/7/1/037.
Kebiche Z, Castelain C, and Burghelea T (2013) Experimental investigation of the Rayleigh–Bénard convection in a yield-stress fluid. *Journal of Non-Newtonian Fluid Mechanics* 203: 9–23.
Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283: 1881–1884.
Kelly A and Bercovici D (1997) The clustering of rising diapirs and plume heads. *Geophysical Research Letters* 24: 201–204.
Kenyon PM and Turcotte DL (1983) Convection in a two- layer mantle with a strongly temperature-dependent viscosity. *Journal of Geophysical Research* 88: 6403–6414.
Kerr RC and Lister JR (1987) The spread of subducted lithospheric material along the mid-mantle boundary. *Earth and Planetary Science Letters* 85: 241–247.
Kerr RC and Lister JR (1988) Island arc and mid-ocean ridge volcanism, modelled by diapirism from linear source regions. *Earth and Planetary Science Letters* 88: 143–152.
Kerr RC and Mériaux C (2004) Structure and dynamics of sheared mantle plumes. *Geochemistry, Geophysics, Geosystems* 5: Q12009. http://dx.doi.org/10.1029/2004GC000749.
Kerr RC, Mériaux C, and Lister JR (2008) Effect of thermal diffusion on the stability of strongly tilted mantle plume tails. *Journal of Geophysical Research* 113: B09401. http://dx.doi.org/10.1029/2007JB005510.
Kerr RC and Lister JR (2008) Rise and deflection of mantle plume tails. *Geochemistry, Geophysics, Geosystems* 9: Q10004. http://dx.doi.org/10.1029/2008GC002124.
King EM, Stellmach S, and Buffet B (2013) Scaling behaviour in Rayleigh–Bénard convection with and without rotation. *Journal of Fluid Mechanics* 717: 449–471.
Kimura I, Hyodo T, and Ozawa M (1998) Temperature and velocity measurement of 3D thermal flow using thermosensitive liquid crystals. *Journal of Visualization* 1: 145–152.
Kincaid C and Olson P (1987) An experimental study of subduction and slab migration. *Journal of Geophysical Research* 92: 13832–13840.
Kincaid C, Ito G, and Gable C (1995) Laboratory investigation of the interaction of off-axis mantle plumes and spreading centres. *Nature* 376: 758–761.
Kincaid C, Sparks D, and Detrick R (1996) The relative importance of plate-driven and buoyancy-driven flow at mid-ocean ridge spreading centers. *Journal of Geophysical Research* 101: 16177–16193.
Kincaid C and Griffiths RW (2003) Laboratory models of the thermal evolution of the mantle during rollback subduction. *Nature* 425: 58–62.
Kincaid C and Griffiths RW (2004) Variability in flow and temperatures within mantle subduction zones. *Geochemistry, Geophysics, Geosystems* 5: Q06002. http://dx.doi.org/10.1029/2003GC000666.
Korenaga J and Jordan TH (2003) Physics of multiscale convection in Earth's mantle: Onset of the sublithospheric convection. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB001760.
Kraichman RH (1962) Turbulent thermal convection at arbitrary Prandtl number. *Physics of Fluids* 5: 1374–1389.
Krishnamurti R (1968a) Finite amplitude convection with changing mean temperature. Part 1. Theory. *Journal of Fluid Mechanics* 33: 445–455.
Krishnamurti R (1968b) Finite amplitude convection with changing mean temperature. Part 2. An experimental test of the theory. *Journal of Fluid Mechanics* 33: 457–463.
Krishnamurti R (1970a) On the transition to turbulent convection. Part 1. Transition from two to three dimensional flow. *Journal of Fluid Mechanics* 42: 295–307.
Krishnamurti R (1970b) On the transition to turbulent convection. Part 2. Transition to time-dependent flow. *Journal of Fluid Mechanics* 42: 309–320.
Krishnamurthi R (1979) Theory and experiment in cellular convection. *Proceedings of NATO Advanced Study Institute on Continental Drift and the Mechanism of Plate Tectonics*, pp. 245–257. England: University of Newcastle upon Tyne.
Krishnamurti R and Howard LN (1981) Large scale flow generation in turbulent convection. *Proceedings of the National Academy of Sciences* 78: 1981–1985.
Kulacki FA and Goldstein RJ (1972) Thermal convection in a horizontal fluid layer with volumetric heat sources. *Journal of Fluid Mechanics* 55: 271–287.
Kulacki FA and Nagle ME (1975) Natural convection in a horizontal fluid layer with volumetric heat sources. *Journal of Heat Transfer* 97: 204–211.
Kumagai I (2002) On the anatomy of mantle plumes: Effect of the viscosity ratio on entrainment and stirring. *Earth and Planetary Science Letters* 198: 211–224.
Kumagai I and Kurita K (2000) On the fate of mantle plumes at density interfaces. *Earth and Planetary Science Letters* 179: 63–71.
Kumagai I, Davaille A, and Kurita K (2007) On the fate of mantle thermal plumes at density interface. *Earth and Planetary Science Letters* 254: 180–193.
Kumagai I, Davaille A, Kurita K, and Stutzmann E (2008) Mantle plumes: Thin, fat, successful, or failing? Constraints to explain hot spot volcanism through time and space. *Geophysical Research Letters* 35: L16301.
Labrosse S (2002) Hotspots, mantle plumes and core heat loss. *Earth and Planetary Science Letters* 199: 147–156.
Laudenbach N (2001) *Experimentelle und numerische Untersuchungen zur Ausbreitung von Volumenstrungen in thermishen Plumes*. PhD Dissertation, Gottingen.
Laudenbach N and Christensen UR (2001) An optical method for measuring temperature in laboratory models of mantle plumes. *Geophysical Journal International* 145: 528–534.
Lavieille P, Lemoine F, Lavergne G, and Lebouché M (2001) Evaporating and combusting droplet temperature measurements using two-color laser-induced fluorescence. *Experiments in Fluids* 31: 45–55.
Le Bars M and Davaille A (2002) Stability of thermal convection in two superimposed miscible viscous fluids. *Journal of Fluid Mechanics* 471: 339–363.
Le Bars M and Davaille A (2004a) Large interface deformation in two-layer thermal convection of miscible viscous fluids. *Journal of Fluid Mechanics* 499: 75–110.
Le Bars M and Davaille A (2004b) Whole-layer convection in an heterogeneous planetary mantle. *Journal of Geophysical Research* 109: B07402. http://dx.doi.org/10.1029/2003JB002617.
Lenardic A, Moresi L-N, Jellinek AM, and Manga M (2005) Continental insulation, mantle cooling, and the surface area of oceans and continents. *Earth and Planetary Science Letters* 234: 317–333.
Lenardic A, Moresi L, Jellinek AM, ONeill CJ, Cooper CM, and Lee CT (2011) Continents, supercontinents, mantle thermal mixing, and mantle thermal isolation: Theory, numerical simulations, and laboratory experiments. *Geochemistry, Geophysics, Geosystems* 12: Q10016. http://dx.doi.org/10.1029/2011GC003663.
Leng W and Zhong SJ (2010) Constraints on viscous dissipation of plate bending from compressible mantle convection. *Earth and Planetary Science Letters* 297: 154–164.
Leong CW and Ottino JM (1990) Increase in regularity by polymer addition during chaotic mixing in two-dimensional flows. *Physical Review Letters* 64: 874–877.
L'evy F and Jaupart C (2012) The initiation of subduction by crustal extension at a continental margin. *Geophysical Journal International* 188: 779–797.
Li Z and Ribe NM (2012) Dynamics of free subduction from 3-D boundary-element modeling. *Journal of Geophysical Research* 117: B06408.
Limare A, Tal M, Reitz MD, Lajeunesse E, and Mtivier F (2011) Optical method for measuring bed topography and flow depth in an experimental flume. *Solid Earth* 2: 143–154.
Limare A, Surducan E, Surducan V, et al. (2013) Microwave-based laboratory experiments for internally-heated mantle convection. In: *Proceedings, 1-8, the International Conference Processes in Isotopes and Molecules (PIM 13), September 25–27, Cluj-Napoca, Romania*.
Lister JR (1989) Selective withdrawal from a viscous two-layer system. *Journal of Fluid Mechanics* 198: 231–254.
Lister JR and Kerr RC (1989a) The effect of geometry on the gravitational instability of a buoyant region of viscous fluid. *Journal of Fluid Mechanics* 202: 231–254.
Lister JR and Kerr RC (1989b) The propagation of two-dimensional and axisymmetric viscous gravity currents along a fluid interface. *Journal of Fluid Mechanics* 203: 215–249.
Lister JR, Kerr RC, Russell NJ, and Crosby A (2011) Rayleigh–Taylor instability of an inclined buoyant viscous cylinder. *Journal of Fluid Mechanics* 671: 313–338.
Lewis WT, Incropera FP, and Viskanta R (1982) Interferometric study of stable salinity gradients heated from below or cooled from above. *Journal of Fluid Mechanics* 116: 411–430.
Lithgow-Bertelloni C, Richards MA, Conrad CP, and Griffiths RW (2001) Plume generation in natural thermal convection at high Rayleigh and Prandtl numbers. *Journal of Fluid Mechanics* 434: 1–21.

Lohmann AW and Silva DE (1972) A Talbot interferometer with circular gratings. *Optics Communications* 4: 326–328.

Lohmann FC, Hort M, and Phipps Morgan J (2009) Flood basalts and ocean island basalts: A deep source or shallow entrainment ? *Earth and Planetary Science Letters* 284: 553–563.

Lohse D and Xia K-Q (2010) Small-scale properties of turbulent Rayleigh–Bénard convection. *Annual Review of Fluid Mechanics* 42: 335–364.

Long MD, Till CB, Druken KA, et al. (2012) Mantle dynamics beneath the Pacific Northwest and the generation of voluminous back-arc volcanism. *Geochemistry, Geophysics, Geosystems* 13: Q0AN01. http://dx.doi.org/10.1029/2012GC004189.

Loper DE and Stacey FD (1983) The dynamical and thermal structure of deep mantle plumes. *Physics of the Earth and Planetary Interiors* 33: 305–317.

Loper DE and Eltayeb IA (1986) On the stability of the D00 layer. *Geophysical & Astrophysical Fluid Dynamics* 36: 229–255.

Loper DE and McCartney K (1986) Mantle plumes and the periodicity of magnetic field reversals. *Geophysical Research Letters* 13: 1525–1528.

Loper DE, McCartney K, and Buzyna G (1988) A model of correlated episodicity in magnetic-field reversals, climate and mass extinctions. *Journal of Geology* 88: 1–14.

Mach E (1873) *Optish-akustusche*. Prague: Calve.

McNamara AK and Zhong S (2004a) Thermochemical structures within a spherical mantle: Superplumes or piles? *Journal of Geophysical Research* 109: B07402. http://dx.doi.org/10.1029/2003JB002847.

McNamara AK and Zhong S (2004b) The influence of thermochemical convection on the fixity of mantle plumes. *Earth and Planetary Science Letters* 222: 485–500.

McNamara AK and Zhong S (2005) Thermochemical structures beneath Africa and the Pacific Ocean. *Nature* 437: 1136–1139.

Malkus WVR (1954) Discrete transitions in turbulent convection. *Proceedings of the Royal Society A* 225: 185–195.

Manga M (1997) Interactions between mantle diapirs. *Geophysical Research Letters* 24: 1871–1874.

Manga M, Stone HA, and OConnell RJ (1993) The interaction of plume heads with compositional discontinuities in the Earth's mantle. *Journal of Geophysical Research* 98: 19979–19990.

Manga M and Weeraratne D (1999) Experimental study of non-Boussinesq Rayleigh–Bénard convection at high Rayleigh and Prandtl numbers. *Physics of Fluids* 10: 2969–2976.

Manga M, Weeraratne D, and Morris SJS (2001) Boundary-layer thickness and instabilities in Bénard convection of a liquid with a temperature-dependent viscosity. *Physics of Fluids* 13: 802–805.

Manga M and Sinton A (2004) Formation of bands, ridges and grooves on Europa by cyclic deformation: Insights from analogue wax experiments. *Journal of Geophysical Research* 109: E09001. http://dx.doi.org/10.1029/2004JE002249.

Maqua C, Castanet G, Lemoine F, Dou N, and Lavergne G (2006) Temperature measurements of binary droplets using three-color laser-induced fluorescence. *Experiments in Fluids* 40: 786–797.

Marsh BD (1979) Island arc development: Some observations and speculations. *Journal of Geology* 87: 687–713.

Marsh BD and Carmichael ISE (1974) Benioff zone magmatism. *Journal of Geophysical Research* 79: 1196–1206.

Martinod J and Davy P (1994) Periodic instabilities during compression of the lithosphere 2. Analogue experiments. *Journal of Geophysical Research* 99: 12057–12070.

Martinod J, Funiciello F, Faccenna C, Labanieh S, and Regard V (2005) Dynamical effects of subducting ridges: Insights from 3-D laboratory models. *Geophysical Journal International* 163: 1137–1150.

Massmeyer A, Di Giuseppe E, Davaille A, Rolf T, and Tackley PJ (2013) Numerical simulation of thermal plumes in a Herschel–Bulkley fluid. *Journal of Non-Newtonian Fluid Mechanics* 195: 32–45.

Matsumoto N, Namiki A, and Sumita I (2006) Influence of a basal thermal anomaly on mantle convection. *Physics of the Earth and Planetary Interiors* 157: 208–222.

Maxwell JC (1873) Double refraction of viscous fluids in motion. *Proceedings of the Royal Society of London, Series A* 22: 46–47.

Merzkirch W (1987) *Flow Visualization*, 2nd edn. Orlando, FL: Academic Press.

Merzkirch W (1989) Streaming birefringence. In: Yang W-J (ed.) *Handbook of Flow Visualization*, pp. 177–180. London: Taylor & Francis.

Merzkirch W (1993) Interferometric methods. *Lecture Series – van Kareman Institute for Fluid Dynamics*, vol. 9, pp. 1–96.

Merzkirch W (1995) Density-sensitive whole-field flow measurement by optical speckle photography. *Experimental Thermal and Fluid Science* 10: 435–443.

Merzkirch W, Vitkin D, and Xiong W (1998) Quantitative flow visualization. *Meccanica* 33: 503–516.

Merzkirch W (2000) Non-intrusive measurement techniques. *Lecture Series, van Karman Institute for Fluid Dynamics*, vol. 6, pp. 1–37.

Meyer C and Schellart WP (2013) Three-dimensional dynamic models of subducting plate-overriding plate-upper mantle interaction. *Journal of Geophysical Research, Solid Earth* 118. http://dx.doi.org/10.1002/jgrb.50078.

Milelli L, Fourel L, and Jaupart C (2012) A lithospheric instability origin for the Cameroon Volcanic Line. *Earth and Planetary Science Letters* 335: 80–87.

Moisy F, Rabaud M, and Salsac K (2009) A synthetic schlieren method for the measurement of the topography of a liquid interface. *Experiments in Fluids* 46: 1021–1036.

Morgan WJ (1971) Convection plumes in the lower mantle. *Nature* 230: 42–43.

Mohr PA and Wood CA (1976) Volcano spacing and lithospheric attenuation in the eastern rift of Africa. *Earth and Planetary Science Letters* 33: 126–144.

Montagner J-P and Tanimoto T (1990) Global anisotropy in the upper mantle-inferred from the regionalization of phase velocities. *Journal of Geophysical Research* 95: 4797–4819.

Montague LN and Kellogg LH (2000) 2000 Numerical models of a dense layer at the base of the mantle and implications for the geodynamics of D. *Journal of Geophysical Research* 105: 11101–11114.

Moore WB, Schubert G, and Tackley P (1998) Three-dimensional simulations of plume–lithosphere interaction at the Hawaiian swell. *Science* 279: 1008–1011.

Moorbath S and Taylor PN (1981) Isotopic evidence for continental growth in the Precambrian. In: Krner A (ed.) *Precambrian Plate Tectonics*, p. 491. Amsterdam: Elsevier.

Moresi L and Solomatov V (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic styles of the Earth and Venus. *Geophysical Journal International* 133: 669–682.

Morris S (1982) The effects of a strongly temperature-dependent viscosity on slow flow past a hot sphere. *Journal of Fluid Mechanics* 124: 1–25.

Morris S and Canright DR (1984) A boundary-layer analysis of Bénard convection with strongly temperature-dependent viscosity. *Physics of the Earth and Planetary Interiors* 36: 355–373.

Morris SW, Bodenschatz E, Cannell DS, and Ahlers G (1993) Spiral defect chaos in large aspect ratio Rayleigh–Bénard convection. *Physical Review Letters* 71: 20–26.

Moses E, Zocchi G, and Libchaber A (1993) An experimental study of laminar plumes. *Journal of Fluid Mechanics* 251: 581–601.

Nagata S (1975) *Mixing: Principles and Applications*. New York: Halsted Press.

Namiki A (2003) Can the mantle entrain D? *Journal of Geophysical Research* 108: 2487. http://dx.doi.org/10.1029/2002JB002315.

Namiki A and Kurita K (1999) The influence of boundary heterogeneity in experimental models of mantle convection. *Geophysical Research Letters* 26: 1929–1932.

Namiki A and Kurita K (2001) The influence of boundary heterogeneity in experimental models of mantle convection with internal heat sources. *Physics of the Earth and Planetary Interiors* 128: 195–205.

Namiki A and Kurita K (2002) Rayleigh–Bénard convection with an inclined upper boundary. *Physical Review E* 65: 056301-1-10.

Namiki A and Kurita K (2003) Heat transfer and interfacial temperature of the two-layered convection: Implications for the D″-mantle coupling. *Geophysical Research Letters* 30: 1108. http://dx.doi.org/10.1029/2002GL015809.

Namiki A, Sueyoshi K, and Takeuchi N (2013) Can a sheet-like low-velocity region form an elongated Large Igneous Province? *Geochemistry, Geophysics, Geosystems* 14: 3053–3066.

Nataf H-C, Froidevaux C, Levrat JL, and Rabinowicz M (1981) Laboratory convection experiments: Effect of lateral cooling and generation of instabilities in the horizontal boundary layers. *Journal of Geophysical Research* 86: 6143–6154.

Nataf H-C, Hager BH, and Scott RF (1984) Convection experiments in a centrifuge and the generation of plumes in a very viscous fluid. *Annals of Geophysics* 2: 303–310.

Nataf H-C, Moreno S, and Cardin P (1988) What is responsible for thermal coupling in layered convection? *Journal of Physics France* 49: 1707–1714.

Nataf H-C (1991) Mantle convection, plates and hotspots. *Tectonophysics* 187: 361–377.

Neavel KE and Johnson AM (1991) Entrainment in compositionally buoyant plumes. *Tectonophysics* 200: 1–15.

Nettleton LL (1934) Fluid mechanics of salt domes. *American Association of Petroleum Geologists Bulletin* 18: 1175–1204.

Nettleton LL (1943) Recent experimental and geophysical evidence of mechanics of salt-dome formation. *American Association of Petroleum Geologists Bulletin* 27: 51–63.

Niemela JJ, Skrbek L, Sreenivasan KR, and Donelly RJ (2000) Turbulent convection at very high Rayleigh numbers. *Nature* 404: 837–840.

Nishimura CE and Forsyth DL (1988) Rayleigh wave phase velocities in the Pacific with implications for azimuthal anisotropy and lateral heterogeneities. *Geophysical Journal International* 94: 479–501.
Oblath N, Daniels K, and Bodenschatz E (2004) Spreading rate dependent morphology of mid-ocean-ridge-like dynamics. *Physical Review Letters* (submitted).
Oertel H and Buhler K (1978) A special differential interferometer used for heat convection investigations. *International Journal of Heat and Mass Transfer* 21: 1111–1115.
Ogawa M, Schubert G, and Zebib A (1991) Numerical simulations of three dimensional thermal convection in a fluid with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 233: 299–328.
Ogawa M (2003) Plate-like regime of a numerically modeled thermal convection in a fluid with temperature-, pressure-, and stress-history-dependent viscosity. *Journal of Geophysical Research* 108: 2067. http://dx.doi.org/10.1029/2000JB000069.
Oldenburg DW and Brune JN (1972) Ridge transform fault spreading patterns in freezing wax. *Science* 178: 301–304.
Oldenburg DW and Brune JN (1975) Explanation for the orthogonality of ocean ridges and transform faults. *Journal of Geophysical Research* 80: 2575–2585.
Olson P (1984) An experimental approach to thermal convection in a two-layered mantle. *Journal of Geophysical Research* 89: 11293–11301.
Olson P and Singer H (1985) Creeping plumes. *Journal of Fluid Mechanics* 158: 511–531.
Olson P and Nam IS (1986) Formation of seafloor swells by mantle plumes. *Journal of Geophysical Research* 91: 7181–7191.
Olson P and Christensen UR (1986) Solitary wave propagation in a fluid conduit within a viscous matrix. *Journal of Geophysical Research* 91: 6367–6374.
Olson P, Schubert G, Anderson C, and Goldman P (1988) Plume formation and lithosphere erosion: A comparison of laboratory and numerical experiments. *Journal of Geophysical Research* 93: 15065–15084.
Olson P (1990) Hot spots, swells and mantle plumes. In: Ryan MP (ed.) *Magma Transport and Storage*, pp. 33–51. New York: Wiley.
Olson P and Kincaid C (1991) Experiments on the interaction of thermal convection and compositional layering at the base of the mantle. *Journal of Geophysical Research* 96: 4347–4354.
Olson P, Schubert G, and Anderson C (1993) Structure of axisymmetric mantle plumes. *Journal of Geophysical Research* 98: 6829–6844.
O'Neill C, Lenardic A, Moresi L, Torsvik TH, and Lee C-TA (2007) Episodic Precambrian subduction. *Earth and Planetary Science Letters* 262: 552–562.
Ottino JM, Leong CW, Rising H, and Swanson PD (1988) Morphological structures produced by mixing in chaotic flows. *Nature* 333: 419–426.
Ottino JM (1989) *The Kinematics of Mixing: Stretching, Chaos, and Transport*, p. 364. Cambridge: Cambridge University Press.
Park H, Dabiri D, and Gharib M (2000) Digital particle image velocimetry/thermometry and application to the wake of a heated circular cylinder. *Experiments in Fluids* 30: 327–338.
Parker TJ and McDowell AN (1951) Scale models as a guide to interpretation of salt-dome faulting. *American Association of Petroleum Geologists Bulletin* 35: 2076–2086.
Parker RL and Oldenburg DW (1973) Thermal model of ocean ridges. *Nature Physical Science* 242: 137–139.
Parmentier EM, Turcotte DL, and Torrance KE (1975) Numerical experiments on the structure of mantle plumes. *Journal of Geophysical Research* 80: 4417–4424.
Parmentier EM and Sotin C (2000) Three-dimensional numerical experiments on thermal convection in a very viscous fluid: Implications for the dynamics of a thermal boundary layer at high Rayleigh number. *Physics of Fluids* 12: 609–617.
Parsons B and McKenzie DP (1978) Mantle convection and the thermal structure of the plates. *Journal of Geophysical Research* 82: 4485–4495.
Parsons B and Sclater JC (1977) An analysis of the variations of ocean floor bathymetry and heat flow with age. *Journal of Geophysical Research* 82: 803–827.
Pearson JRA (1958) On convection cells induced by surface tension. *Journal of Fluid Mechanics* 4: 489–500.
Piau J (2007) Carbopol gels: Elastoviscoplastic and slippery glasses made of individual swollen sponges: Meso- and macroscopic properties, constitutive equations and scaling laws. *Journal of Non-Newtonian Fluid Mechanics* 144: 1–29.
Pindera JT and Krishnamurthy AR (1978) Characteristic relations of flow birefringence: Part 2. Relations in Scattered radiations. *Experimental Mechanics* 18: 41–48.
Pinet N and Cobbold PR (1992) Experimental insights into the partitioning of motion within zones of oblique subduction. *Tectonophysics* 206: 371–388.
Plapp BB (1997) *Spiral-pattern formation in Rayleigh–Bénard convection*. PhD Thesis. Cornell University, Ithaca, NY.
Plapp BB, Egolf DA, Bodenschatz E, and Pesch W (1998) Dynamics and selection of giant spirals in Rayleigh–Bénard convection. *Physical Review Letters* 81: 5334–5337.
Pottebaum TS and Gharib M (2004) The pinch-off process in a starting buoyant plume. *Experiments in Fluids* 37: 87–94.
Pouliquen O and Forterre Y (2002) Friction law for dense granular flows: Application to the motion of a mass down a rough inclined plane. *Journal of Fluid Mechanics* 453: 131–151.
Pretzel G, Jger H, and Neger T (1993) High-accuracy differential interferometry for the investigation of phase objects. *Measurement Science and Technology* 4: 649–658.
Pubellier M and Cobbold PR (1996) Analogue models for the transpressional docking of volcanic arcs in the western Pacific. *Tectonophysics* 253: 33–52.
Puthenveettil BA and Arakeri JH (2005) Plume structure in high-Rayleigh-number convection. *Journal of Fluid Mechanics* 542: 217–249.
Qi JA, Leung CW, Wong WO, and Probert SD (2006) Temperature-field measurements of a premixed butane/air circular impinging-flame using reference-beam interferometry. *Applied Energy* 83: 1307–1316.
Raffel M, Willert C, and Kompenhans J (1998) *Particle Image Velocimetry*. Berlin: Springer.
Ragnarsson R, Ford JL, Santangelo CD, and Bodenschatz E (1996) Rifts in spreading wax layers. *Physical Review Letters* 76: 3456–3459.
Ramberg H (1967) Model experimentation of the effect of gravity on tectonic processes. *Geophysical Journal of the Royal Astronomical Society* 14: 307–329.
Ramberg H (1968) Instability of layered systems in the field of gravity. *Physics of the Earth and Planetary Interiors* 1: 427–447.
Ramberg H (1972) Mantle diapirism and its tectonic and magmatic consequences. *Physics of the Earth and Planetary Interiors* 4: 45–60.
Ramberg H (1975) Particle paths, displacement and progressive strain applicable to rocks. *Tectonophysics* 28: 1–35.
Ramberg H and Sjostrom H (1973) Experimental geodynamical models relating to continental drift and orogenesis. *Tectonophysics* 19: 105–132.
Ramberg H and Ghosh SK (1977) Rotation and strain of linear and planar structures in three-dimensional progressive deformation. *Tectonophysics* 40: 309–337.
Rasenat S, Busse FH, and Rehberg I (1989) A theoretical and experimental study of double-layer convection. *Journal of Fluid Mechanics* 199: 519–540.
Ratcliff JT, Tackley PJ, Schubert G, and Zebib A (1997) Transitions in thermal convection with strongly variable viscosity. *Physics of the Earth and Planetary Interiors* 102: 201–212.
Rayleigh Lord (1883) Investigations of the character of an incompressible heavy fluid of variable density. *Proceedings of the London Mathematical Society* 14: 170–177.
Rayleigh Lord (1916) On convection currents in a horizontal layer of fluid, when the higher temperature is on the underside. *Philosophical Magazine* 32: 529–546.
Regard V, Faccenna C, Bellier O, and Martinod J (2008) Laboratory experiments of slab break-off and slab dip reversal: Insight into the Alpine Oligocene reorganization. *Terra Nova* 20: 17. http://dx.doi.org/10.1111/j.1365- 3121.2008.00815.x.
Renardy M and Renardy Y (1985) Perturbation analysis of steady and oscillatory onset in a Bénard problem with two similar liquids. *Physics of Fluids* 28: 2699–2708.
Renardy Y and Joseph DD (1985a) Oscillatory instability in a Bénard problem of two fluids. *Physics of Fluids* 28: 788–793.
Renardy Y and Joseph DD (1985b) Perturbation analysis of steady and oscillatory onset in a Bénard problem with two similar liquids. *Physics of Fluids* 28: 2699–2708.
Reymer A and Schubert G (1986) Rapid growth of some major segments of continental crust. *Geology* 14: 299–302.
Reynolds O (1894) Study of fluid motion by means of colored bands. *Nature* 50: 161–164.
Rhee H, Koseff J, and Street R (1984) Flow visualization of a recirculating flow by rheoscopic and liquid crystal techniques. *Experiments in Fluids* 2: 57–64.
Ribe NM (1998) Spouting and planform selection in the Rayleigh–Taylor instability of miscible viscous fluids. *Journal of Fluid Mechanics* 234: 315–336.
Ribe NM and Christensen U (1994) Three-dimensional modeling of plume–lithosphere interaction. *Journal of Geophysical Research* 99: 669–682.
Ribe NM and de Valpine DP (1994) The global hotspot distribution and instability of D. *Geophysical Research Letters* 21: 1507–1510.
Ribe NM, Christensen UR, and Theissing J (1995) The dynamics of plume–ridge interaction, 1: Ridge-centered plumes. *Earth and Planetary Science Letters* 134: 155–168.
Ribe NM and Christensen U (1999) The dynamical origin of Hawaiian volcanism. *Earth and Planetary Science Letters* 171: 517–531.
Ribe NM (2003) Periodic folding of viscous sheets. *Physical Review E* 68: 036305.

Ribe NM, Stutzmann E, Ren Y, and van der Hilst R (2007) Buckling instabilities of subducted lithosphere beneath the transition zone. *Earth and Planetary Science Letters* 254: 173–179.

Ribe NM (2010) Bending mechanics and mode selection in free subduction: a thin-sheet analysis. *Geophysical Journal International* 180: 559–576.

Ribe NM and Davaille A (2013) Dynamical similarity and density (non-) proportionality in experimental tectonics. *Tectonophysics* 608: 1371–1379.

Ricard Y, Vigny C, and Froidevaux C (1989) Mantle heterogeneities, geoid, and plate motion: A Monte Carlo inversion. *Journal of Geophysical Research* 94: 13739–13754.

Richards MA and Griffiths RW (1988) Deflection of plumes by mantle shear flow: Experimental results and a simple theory. *Geophysical Journal* 94: 367–376.

Richards MA, Duncan RA, and Courtillot VE (1989) Flood basalts and hotspot tracks: Plume heads and tails. *Science* 246: 103–107.

Richter FM (1973) Convection and the large-scale circulation of the mantle. *Journal of Geophysical Research* 78: 8735–8745.

Richter FM (1978) Experiments on the stability of convection rolls in fluids whose viscosity depends on temperature. *Journal of Fluid Mechanics* 89: 553.

Richter FM and Johnson CE (1974) Stability of a chemically layered mantle. *Journal of Geophysical Research* 79: 1635–1639.

Richter FM and Parsons B (1975) On the interaction of two scales of convection in the mantle. *Journal of Geophysical Research* 80: 2529–2541.

Richter FM and McKenzie DP (1978) Simple plate models of mantle convection. *Journal of Geophysics* 44: 441.

Richter FM and McKenzie DP (1981) On some consequences and possible causes of layered convection. *Journal of Geophysical Research* 86: 6124–6133.

Richter FM, Nataf H-C, and Daly SF (1983) Heat transfer and horizontally averaged temperature of convection with large viscosity variations. *Journal of Fluid Mechanics* 129: 173–192.

Roche P-E, Gauthier F, Kaiser R, and Salort J (2010) On the triggering of the Ultimate Regime of convection. *New Journal of Physics* 12(2010): 085014.

Rogers MC and Morris SW (2009) Natural versus forced convection in laminar starting plumes. *Physics of Fluids* 21: 083601.

Rossby HT (1965) On thermal convection driven by non-uniform heating form below: An experimental study. *Deep-Sea Research* 12: 9–16.

Rossby HT (1969) A study of Bénard convection with and without convection. *Journal of Fluid Mechanics* 36: 309–335.

Rouzo S, Rabinowicz M, and Briais A (1995) Segmentation of mid-ocean ridges with an axial valley induced by small-scale convection. *Nature* 374: 795–798.

Russo RM and Silver PG (1994) Trench-parallel flow beneath the Nazca plate from seismic anisotropy. *Science* 263: 1105–1111.

Samuel H and Farnetani CG (2002) Thermochemical convection and helium concentrations in mantle plumes. *Earth and Planetary Science Letters* 207: 39–56.

Samuel H and Bercovici D (2006) Oscillating and stagnating plumes in the Earth's lower mantle. *Earth and Planetary Science Letters* 248: 90–105.

Schaeffer N and Manga M (2001) Interaction of rising and sinking mantle plumes. *Geophysical Research Letters* 21: 765–768.

Schellart WP (2002) Analogue modelling of large-scale tectonic processes. *Journal of the Virtual Explorer* 7: 1–6.

Schellart WP (2004a) Kinematics of subduction and subduction-induced flow in the upper mantle. *Journal of Geophysical Research* 109: B07401. http://dx.doi.org/10.1029/2004JB002970.

Schellart WP (2004b) Quantifying the net slab pull force as a driving mechanism for plate tectonics. *Geophysical Research Letters* 31: L07611. http://dx.doi.org/10.1029/2004GL019528.

Schellart WP (2005) Influence of the subducting plate velocity on the geometry of the slab and migration of the subduction hinge. *Earth and Planetary Science Letters* 231: 197–219.

Schellart WP, Freeman J, Stegman DR, Moresi L, and May D (2007) Evolution and diversity of subduction zones controlled by slab width. *Nature* 446(7133): 308–311.

Schellart WP (2008) Kinematics and flow patterns in deep mantle and upper mantle subduction models: Influence of the mantle depth and slab to mantle viscosity ratio. *Geochemistry, Geophysics, Geosystems* 9: Q03014. http://dx.doi.org/10.1029/2007GC001656.

Schellart WP (2010) Evolution of subduction zone curvature and its dependence on the trench velocity and the slab to upper mantle viscosity ratio. *Journal of Geophysical Research* 115: B11406. http://dx.doi.org/10.1029/2009JB006643.

Schilling J-G (1991) Fluxes and excess temperatures of mantle plumes inferred from their interaction with migrating mid-ocean ridges. *Nature* 352: 397–403.

Schmitz E and Merzkirch W (1981) A test fluid for simulating blood flows. *Experiments in Fluids* 2: 103–104.

Schmeling H (1988) Numerical models of Rayleigh–Taylor instabilities superimposed upon convection. *Bulletin of the Geological Institutions of the University of Uppsala* 14: 95–109.

Schmeling H, et al. (2008) A benchmark comparison of spontaneous sub-duction models – Towards a free surface. *Physics of the Earth and Planetary Interiors* 171: 198–223.

Schmidt RJ and Milverton SW (1935) On the instability of a fluid when heated from below. *Proceedings of the Royal Society A* 152: 586–594.

Schouten H, Klitgord KD, and Gallow DG (1993) Edge-driven microplate kinematics. *Journal of Geophysical Research* 98(B4): 6689–6701.

Schubert G, Olson P, and Anderson Goldman P (1989) Solitary waves in mantle plumes. *Journal of Geophysical Research* 94: 9523–9532.

Schubert G, Masters G, Olson P, and Tackley P (2004) Superplumes or plume clusters? *Physics of the Earth and Planetary Interiors* 146: 147–162.

Schwiderski EW and Schwab HJA (1971) Convection experiments with electrolytically heated fluid layers. *Journal of Fluid Mechanics* 48: 703–717.

Scott DR, Stevenson DJ, and Whitehead JA (1987) Observations of solitary waves in a viscously deformable pipe. *Nature* 319: 759–761.

Selig F (1965) A theoretical prediction if salt-dome patterns. *Geophysics* 30: 633–645.

Seon T, Hulin JP, Salin D, Perrin B, and Hinch EJ (2006) Laser-induced fluorescence measurements of buoyancy driven mixing in tilted tubes. *Physics of Fluids* 18: 041701.

Sernas V and Fletcher LS (1970) A schlieren interferometer method for heat transfer studies. *Journal of Heat Transfer* 92: 202–204.

Settles GS (2001) *Schlieren and Shadowgraph Techniques; Visualizing Phenomena in Transparent Media*. Heidelberg: Springer-Verlag.

Shemenda AI (1981) The conditions and some results from physical modeling of lithospheric deformation (in Russian). In: *Problems of Theoretical Geodynamics and the Tectonics of Lithosphere Plates*, pp. 146–154. Moscow: Institut Okeanologii Akad. Nauk SSSR.

Shemenda AI (1992) Horizontal lithosphere compression and subduction: Constraints provided by physical modeling. *Journal of Geophysical Research* 97: 1097–1116.

Shemenda AI (1993) Subduction of the lithosphere and back arc dynamics: Insights from physical modeling. *Journal of Geophysical Research* 98: 16167–16185.

Shlien DJ (1976) Some laminar thermal and plume experiments. *Physics of Fluids* 19: 1089–1098.

Shlien DJ and Thompson DJ (1975) Some experiments on the motion of an isolated laminar thermal. *Journal of Fluid Mechanics* 72: 35–47.

Shlien DJ and Boxman RL (1979) Temperature field measurement of an axisymmetric laminar plume. *Physics of Fluids* 22: 631–634.

Shlien DJ and Brosh A (1979) Velocity field measurements of a laminar thermal. *Physics of Fluids* 22: 1044–1053.

Shlien DJ and Boxman RL (1981) Laminar starting plume temperature field measurement. *International Journal of Heat and Mass Transfer* 24: 919–930.

Schouten J, Klitgord KD, and Whitehead JA (1985) Segmentation of mid-ocean ridges. *Nature* 317: 225–229.

Shew WL, Gasteuil Y, Gibert M, Metz P, and Pinton J-F (2007) Instrumented tracer for Lagrangian measurements in Rayleigh–Bénard convection. *Review of Scientific Instruments* 78: 065105.

Siggia ED (1994) High Rayleigh number convection. *Annual Review of Fluid Mechanics* 26: 137–168.

Sigurdsson H and Sparks SRJ (1978) Lateral magma flow within rifted Icelandic crust. *Nature* 274: 126–130.

Silveston PL (1958) Wärmedurchgang in waagerechten Flüssigkeitschichten. *Forschung auf dem Gebiet des Ingenieurwesens A* 24: 59–69.

Singh S, Carton H, Chauhan AS, et al. (2011) Extremely thin crust in the Indian Ocean possibly resulting from plume–ridge Interaction. *Geophysical Journal International* 184: 29–42.

Skilbeck JN and Whitehead JA (1978) Formation of discrete islands in linear islands chains. *Nature* 272: 499–501.

Sleep NH (1988) Gradual entrainment of a chemical layer at the base of the mantle by overlying convection. *Geophysical Journal* 95: 437–447.

Sleep NH (1990) Hotspots and mantle plumes: Some phenomenology. *Journal of Geophysical Research* 95: 6715–6736.

Sleep NH (1994) Lithospheric thinning by midplate mantle plumes and the thermal history of hot plume material ponded at sublithospheric depths. *Journal of Geophysical Research* 99: 9327–9343.

Small RD, Sernas VA, and Page RH (1972) *Applied Optics* 11: 858–862.

Solomatov S (1995) Scaling of temperature- and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.

Solomatov VS and Moresi LN (2000) Scaling of time-dependent stagnant lid convection: Application to small-scale convection on Earth and other terrestrial planets. *Journal of Geophysical Research* 105: 21785–21817.
Solomon TH and Gollub JP (1988a) Passive transport in steady Rayleigh–Bénard convection. *Physics of Fluids* 31: 1372–1379.
Solomon TH and Gollub JP (1988b) Chaotic particle transport in time-dependent Rayleigh–Bénard convection. *Physical Review A* 38: 6280–6286.
Solomon T and Gollub J (1990) Sheared boundary layers in turbulent Rayleigh–Bénard convection. *Physical Review Letters* 64: 2382–2385.
Solomon T and Gollub J (1991) Thermal boundary layers and heat flux in turbulent convection: The role of recirculating flows. *Physical Review A* 43: 6683–6693.
Solomon TH, Tomas S, and Warner JL (1996) The role of lobes in chaotic mixing of miscible and immiscible impurities. *Physical Review Letters* 77: 2682–2685.
Solomon TH and Mezic I (2003) Uniform resonant chaotic mixing in fluid flows. *Nature* 425: 376–380.
Sotin C and Parmentier EM (1989) On the stability of a fluid layer containing a univariant phase transition: Application to planetary interiors. *Physics of the Earth and Planetary Interiors* 55: 10–25.
Sparks DW and Parmentier EM (1993) The structure of three-dimensional convection beneath oceanic spreading centers. *Geophysical Journal International* 112: 81–91.
Sparrow EM, Husar RB, and Goldstein RJ (1970) Observations and other characteristics of thermals. *Journal of Fluid Mechanics* 41: 793–800.
Stacey FD and Loper DE (1983) The thermal boundary layer interpretation of D" and its role as a plume source. *Physics of the Earth and Planetary Interiors* 33: 45–55.
Stacey FD and Loper DE (1984) Thermal histories of the core and mantle. *Physics of the Earth and Planetary Interiors* 36: 99–115.
Stegman DR, Freeman J, Schellart WP, Moresi L, and May D (2006) Influence of trench width on subduction hinge retreat rates in 3-D models of slab rollback. *Geochemistry, Geophysics, Geosystems* 7: Q03012. http://dx.doi.org/10.1029/2005GC001056.
Stegman DR, Farrington R, Capitanio FA, and Schellart WP (2010) A regime diagram for subduction styles from 3-D numerical models of free subduction. *Tectonophysics* 483: 29–45.
Stein C, Schmalz G, and Hansen U (2004) The effects of rheological parameters on plate behaviour in a self-consistent model of mantle convection. *Physics of the Earth and Planetary Interiors* 142: 225–255.
Stengel KC, Oliver DS, and Booker JR (1982) Onset of convection in a variable-viscosity fluid. *Journal of Fluid Mechanics* 120: 411–431.
Stein M and Hofmann AW (1994) Mantle plumes and episodic crustal growth. *Nature* 372: 63–68.
Stevens RJAM, van der Poel EP, Grossmann S, and Lohse D (2013) The unifying theory of scaling in thermal convection: The updated prefactors. *Journal of Fluid Mechanics* 730: 295.
Swanson WM and Green RL (1969) Colloidal suspension properties of Milling Yellow dye. *Journal of Colloid and Interface Science* 29: 161–163.
Tackley PJ, Stevenson DJ, Glatzmaier GA, and Schubert G (1994) Effects of multiple phase transitions in a 3-dimensional spherical model of convection in Earth's mantle. *Journal of Geophysical Research* 99: 15887–15901.
Tackley PJ (1996) Effects of strongly variable viscosity on three-dimensional compressible convection in planetary mantles. *Journal of Geophysical Research* 101: 3311–3332.
Tackley PJ (1998) Three-dimensional simulations of mantle convection with a thermo-chemical basal boundary layer: D? In: Gurnis M, Wysession M, Knittle E, and Buffett B (eds.) *The Core–Mantle Boundary Region*, pp. 231–253. Washington, DC: American Geophysical Union.
Tackley PJ (2000a) Mantle convection and Plate Tectonics: Toward an integrated physical and chemical theory. *Science* 288: 2002–2007.
Tackley PJ (2000b) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations 1. Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* 1: 2000Gc000041.
Tackley PJ (2002) Strong heterogeneity caused by deep mantle layering. *Geochemistry, Geophysics, Geosystems* 3: 1–22. http://dx.doi.org/10.1029/2001GC000167.
Tait S and Jaupart C (1989) Compositional convection in viscous melts. *Nature* 338: 571–574.
Takahashi J, Tasaka Y, Murai Y, Takeda Y, and Yanagisawa T (2010) Experimental study of cell pattern formation induced by internal heat sources in a horizontal fluid layer. *International Journal of Heat and Mass Transfer* 53(7–8): 1483–1490.
Tanny J and Shlien DJ (1985) Velocity field measurements of a laminar starting plume. *Physics of Fluids* 28: 1027–1032.
Tamai N and Asaeda T (1984) Sheetlike plumes near a heated bottom plate at large Rayleigh number. *Journal of Geophysical Research* 89: 727–734.
Tan E, Gurnis M, and Han L (2002) Slabs in the lower mantle and their modulation of plume formation. *Geochemistry, Geophysics, Geosystems* 3(11): 1067. http://dx.doi.org/10.1029/2001GC000238.
Tan E and Gurnis M (2005) Metastable superplumes and mantle compressibility. *Geophysical Research Letters* 32. http://dx.doi.org/10.1029/2005GL024190.
Tan E, Leng W, Zhong S, and Gurnis M (2011) On the location of plumes and lateral movement of thermo-chemical structures with high bulk modulus in the 3-D compressible mantle. *Geochemistry, Geophysics, Geosystems* 12: Q07005. http://dx.doi.org/10.1029/ 2011GC003665.
Tanner LH (1966) Some applications of holography in fluid mechanics. *Journal of Scientific Instruments* 43: 81–83.
Tasaka Y, Kudoh Y, Takeda Y, and Yanagisawa T (2005) Experimental investigation of natural convection induced by internal heat generation. *Journal of Physics Conference Series* 14: 168–179.
Taylor G (1950) The instability of liquid surfaces when accelerated in a direction perpendicular to their planes. *Proceedings of the Royal Society of London, Series A* 201: 192–196.
Ten Brink U (1991) Volcano spacing and plate rigidity. *Geology* 19: 397–400.
Tentler T (2003a) Analogue modeling of tension fracture pattern in relation to mid-ocean ridge propagation. *Geophysical Research Letters* 30: 1268.
Tentler T (2003b) Analogue modeling of overlapping spreading centers: Insights into their propagation and coalescence. *Tectonophysics* 376: 99–115.
Tentler T (2007) Focused and diffuse extension in controls of ocean ridge segmentation in analogue models. *Tectonics* 26: TC5008.
Tentler T and Acocella V (2010) How does the initial configuration of oceanic ridge segments affect their interaction? Insights from analogue models. *Journal of Geophysical Research* 115: B01401.
Thompson PF and Tackley PJ (1998) Generation of mega-plumes from the core–mantle boundary in a compressible mantle with temperature-dependent viscosity. *Geophysical Research Letters* 25: 1999–2002.
Thomson J (1882) On a changing tessellated structure in certain liquids. *Proceedings of the Philosophical Society of Glasgow* 13: 464–468.
Torrance KE and Turcotte DL (1971) Thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 47: 113–125.
Townsend AA (1957) Temperature fluctuations over a heated horizontal surface. *Journal of Fluid Mechanics* 5: 209–241.
Tritton DJ and Zarraga MM (1967) Convection in horizontal fluid layers with internal heat generation experiments. *Journal of Fluid Mechanics* 30: 21–31.
Trompert RA and Hansen U (1998a) On the Rayleigh number dependence of convection with a strongly temperature-dependent viscosity. *Physics of Fluids* 10: 351–360.
Trompert RA and Hansen U (1998b) Mantle convection simulations with rheologies that generate plate-like behaviour. *Nature* 395: 686–689.
Tsinober AB, Yahalom Y, and Shlien DJ (1983) A point source of heat in a stable salinity gradient. *Journal of Fluid Mechanics* 135: 199–217.
Turner JS (1973) Convection in the mantle: A laboratory model with temperature-dependent viscosity. *Earth and Planetary Science Letters* 17: 369–374.
Van Keken P (1997) Evolution of starting mantle plumes: A comparison between numerical and laboratory models. *Earth and Planetary Science Letters* 148: 1–11.
van Keken P, King S, Schmeling H, Christensen U, Neumeister D, and Doin M-P (1997) A comparison of methods for the modeling of thermochemical convection. *Journal of Geophysical Research* 102: 22477–22496.
van Keken PE, Currie C, King SD, et al. (2008) A community benchmark for subduction zone modeling. *Physics of the Earth and Planetary Interiors* 171: 187–197.
van Keken PE, Davaille A, and Vatteville J (2013) Dynamics of a laminar plume in a cavity: The influence of boundaries on the steady-state stem structure. *Geochemistry, Geophysics, Geosystems* 14: 158–178. http://dx.doi.org/10.1029/2012GC004383.
Vasquez PA, Perez AT, and Castellanos A (1996) Thermal and electrohydrodynamic plumes. A comparative study. *Physics of Fluids* 8: 2091–2096.
Vatteville J (2004) *Etude exprimentale des panaches: Structure et evolution temporelle.* IPG Paris: Mmoire de DEA, ENS Lyon.
Vatteville J, van Keken PE, Limare A, and Davaille A (2009) Starting laminar plumes: Comparison of laboratory and numerical modeling. *Geochemistry, Geophysics, Geosystems* 10: Q12013. http://dx.doi.org/10.1029/2009GC002739.
Verzicco R (2004) Effect of non-perfect thermal sources in turbulent thermal convection. *Physics of Fluids* 16: 1965–1975.
Vest CM (1975) Interferometry of strongly refracting axisymmetric phase objects. *Applied Optics* 14: 1601–1606.
Villermaux E and Innocenti C (1999) On the geometry of turbulent mixing. *Journal of Fluid Mechanics* 393: 123–147.

Voth GA, Haller G, and Gollub JP (2002) Experimental measurements of stretching fields in fluid mixing. *Physical Review Letters* 88: 254–501.
Weeraratne D and Manga M (1998) Transitions in the style of mantle convection at high Rayleigh numbers. *Earth and Planetary Science Letters* 160: 563–568.
Weijermars R (1986) Flow behavior and physical chemistry of bouncing putties and related polymers in view of tectonics laboratory applications. *Tectonophysics* 124: 325–358.
Weijermars R and Schmeling H (1986) Scaling of Newtonian and non-Newtonian fluid dynamics without inertia for quantitative modelling of rock flow due to gravity (including the concept of rheological similarity). *Physics of the Earth and Planetary Interiors* 43: 316–330.
Weijermars R (1988) New laboratory method for analyzing deformation and displacement in creeping fluid: Examples from stokes flow and a thermal plume. *Journal of Geophysical Research* 93: 2179–2190.
Weijermars R (1989) Experimental pictures of deformation patterns in a possible model of the Earth's interior. *Earth and Planetary Science Letters* 91: 367–373.
Weinstein SA and Olson PL (1989) The proximity of hotspots to convergent and divergent plate boundaries. *Geophysical Research Letters* 16: 433–436.
Weinstein SA and Olson P (1990) Planforms in thermal convection with internal heat sources at large Rayleigh and Prandtl numbers. *Geophysical Research Letters* 17: 239–242.
Weinstein SA and Christensen UR (1991) Convection planforms in a fluid with a temperature-dependent viscosity beneath a stress-free upper boundary. *Geophysical Research Letters* 18: 2035–2038.
Wenzel MJ, Manga M, and Jellinek AM (2004) Tharsis as a consequence of Marsdichotomy and layered mantle. *Geophysical Research Letters* 31: L04702. http://dx.doi.org/10.1029/2003GL019306.
White DB (1988) The planforms and onset of convection with a temperature-dependent viscosity fluid? *Journal of Fluid Mechanics* 191: 247–286.
Whittaker RJ and Lister JR (2006a) Steady axisymmetric creeping plumes above a planar boundary. Part I: a point source. *Journal of Fluid Mechanics* 567: 361–378.
Whittaker RJ and Lister JR (2006b) Steady axisymmetric creeping plumes above a planar boundary. Part II: a distributed source. *Journal of Fluid Mechanics* 567: 379–397.
Whitehead JA (1972) Moving heaters as a model of continental drift. *Physics of the Earth and Planetary Interiors* 18: 199–212.
Whitehead JA and Luther DS (1975) Dynamics of laboratory diapir and plume models. *Journal of Geophysical Research* 80: 705–717.
Whitehead JA (1976) Convection models: Laboratory versus mantle. *Tectonophysics* 35: 215–228.
Whitehead JA and Parsons B (1978) Observations of convection at Rayleigh numbers up to 760000 in a fluid with large Prandtl number. *Geophysical & Astrophysical Fluid Dynamics* 9: 201–217.
Whitehead JA (1982) Instabilities of fluid conduits in a flowing Earth are plates lubricated by the asthenosphere? *Geophysical Journal of the Royal Astronomical Society* 70: 415–433.
Whitehead JA, Dick HHB, and Schouten H (1984) A mechanism for magmatic accretion under spreading centers. *Nature* 312: 146–148.
Whitehead JA (1986) Buoyancy driven instabilities of low viscosity zones as models of magma-rich zones. In: Waff Harve S and Kumazawa Mineo (eds.) *Partial Melting Phenomena in the Earth and Planetary Evolution: special issue*, *Journal of Geophysical Research*, 91, pp. 9303–9314.
Whitehead JA and Helfrich KR (1988) Wave transport of deep mantle material. *Nature* 336: 59–61.
Wilcock WSD and Whitehead JA (1991) The Rayleigh–Taylor instability of an embedded layer of low-viscosity fluid. *Journal of Geophysical Research* 96: 12193–12200.
Willert C and Gharib M (1991) Digital particle image thermometry. *Experiments in Fluids* 10: 181–193.
Willis GE and Deardorff JW (1967) Confirmation and renumbering of the discrete heat flux transitions of Malkus. *Physics of Fluids* 10: 1861–1866.
Wilson TJ (1963) Evidence from islands on the spreading of the ocean floor. *Canadian Journal of Physics* 41: 863–868.
Worster MG (1986) The axisymmetric laminar plume: Asymptotic solution for large Prandtl number. *Studies in Applied Mathematics* 75: 139–152.
Wu B, Conrad CP, Heuret A, Lithgow-Bertelloni A, and Lallemand S (2008) Reconciling strong slab pull and weak plate bending: the plate motion constraint on the strength of mantle slabs. *Earth and Planetary Science Letters* 272: 412–421.
Xi H-D, Lam S, and Xia K-Q (2004) From laminar plumes to organized flows: The onset of large-scale circulation in turbulent thermal convection. *Journal of Fluid Mechanics* 503: 47–56.
Yang W-J (1989) *Handbook of Flow Visualization*. London: Taylor & Francis.
Zaranek SE and Parmentier EM (2004) The onset of convection in fluids with strongly temperature-dependent viscosity cooled from above with implications for planetary lithospheres. *Earth and Planetary Science Letters* 224: 71–386.
Zhang J, Childress S, and Libchaber A (1997) Non-Boussinesq effect: Thermal convection with broken symmetry. *Physics of Fluids* 9: 1034–1042.
Zhang J, Childress S, and Libchaber A (1998) Non-Boussinesq effect: Asymmetric velocity profiles in thermal convection. *Physics of Fluids* 10: 1534–1536.
Zhang J and Libchaber A (2000) Periodic boundary motion in thermal turbulence. *Physical Review Letters* 84: 4361–4364.
Zhong J-Q and Zhang J (2005) Thermal convection with a freely moving top boundary. *Physics of Fluids* 17: 115105.
Zhou Q, Sun C, and Xia K-Q (2007) Morphological evolution of thermal plumes in turbulent Rayleigh–Bénard convection. *Physical Review Letters* 98: 074501.

7.04 Analytical Approaches to Mantle Dynamics

Neil M Ribe, Université Paris-Sud and CNRS, Orsay, France

© 2015 Elsevier B.V. All rights reserved.

7.04.1	**Introduction**	146
7.04.2	**Formulating Geodynamical Model Problems: Three Case Studies**	147
7.04.2.1	Heat Transfer from Mantle Diapirs	147
7.04.2.2	Plume Formation in TBLs	148
7.04.2.3	Plume–Lithosphere Interaction	149
7.04.3	**Dimensional and Scaling Analysis**	150
7.04.3.1	Buckingham's Π-Theorem and Dynamical Similarity	150
7.04.3.2	Nondimensionalization	152
7.04.3.3	Scaling Analysis	153
7.04.4	**Self-Similarity and Intermediate Asymptotics**	154
7.04.4.1	Conductive Heat Transfer	154
7.04.4.2	Self-Similar Buoyant Thermals	155
7.04.4.3	Classification of Self-Similar Solutions	156
7.04.4.4	Intermediate Asymptotics With Respect to Parameters: The R–T Instability	156
7.04.5	**Slow Viscous Flow**	157
7.04.5.1	Basic Equations and Theorems	158
7.04.5.2	Potential Representations for Incompressible Flow	159
7.04.5.2.1	2-D and axisymmetric flows	159
7.04.5.2.2	3-D flows	159
7.04.5.3	Classical Exact Solutions	160
7.04.5.3.1	Steady unidirectional flow	160
7.04.5.3.2	Stokes–Hadamard solution for a sphere	160
7.04.5.3.3	Models for subduction zones and ridges	161
7.04.5.3.4	Viscous eddies	162
7.04.5.4	Superposition and Eigenfunction Expansion Methods	162
7.04.5.4.1	2D flow in Cartesian coordinates	162
7.04.5.4.2	Spherical coordinates	164
7.04.5.4.3	Bispherical coordinates	164
7.04.5.5	The Complex-Variable Method for 2-D Flows	164
7.04.5.6	Singular Solutions and the Boundary Integral Representation	165
7.04.5.6.1	Flow due to point forces	165
7.04.5.6.2	Flow due to point sources	165
7.04.5.6.3	Singular solutions in the presence of a boundary	165
7.04.5.6.4	Boundary-integral representation	166
7.04.5.7	Slender-Body Theory	167
7.04.5.8	Flow Driven by Internal Loads	168
7.04.5.8.1	Wave-domain Green's functions	168
7.04.5.8.2	The propagator matrix method	170
7.04.6	**Elasticity and Viscoelasticity**	171
7.04.6.1	Correspondence Principles	171
7.04.6.2	Surface Loading of a Stratified Elastic Sphere	172
7.04.7	**Boundary-Layer Theory**	173
7.04.7.1	The Boundary-Layer Approximation	173
7.04.7.2	Solution of the Boundary-Layer Equations Using Variable Transformations	173
7.04.7.2.1	The von Mises transformation	174
7.04.7.2.2	Mangler's transformation	174
7.04.7.3	The Method of Matched Asymptotic Expansions	175
7.04.7.4	Boundary Layers with Strongly Variable Viscosity	176
7.04.8	**Long-Wave Theories**	177
7.04.8.1	Lubrication Theory	177
7.04.8.2	Plume–Plate and PRI Models	178
7.04.8.3	Long-Wave Analysis of Buoyant Instability	179
7.04.8.4	Theory of Thin Shells, Plates, and Sheets	180
7.04.8.5	Effective Boundary Conditions from Thin-Layer Flows	182

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00129-9

7.04.8.6 Solitary Waves 182
7.04.9 Hydrodynamic Stability and Thermal Convection 183
7.04.9.1 Rayleigh–Taylor Instability 183
7.04.9.2 Rayleigh–Bénard Convection 184
7.04.9.2.1 Linear stability analysis 185
7.04.9.2.2 Order-parameter equations for finite-amplitude thermal convection 185
7.04.9.2.3 Amplitude equation for convection rolls 186
7.04.9.2.4 Finite-amplitude convection rolls and their stability 187
7.04.9.2.5 Envelope equation for modulated convection rolls 188
7.04.9.2.6 Phase diffusion equation for thermal convection 189
7.04.9.3 Convection at High Rayleigh Number 190
7.04.9.3.1 Scaling analysis 190
7.04.9.3.2 Flow in the isothermal core 190
7.04.9.3.3 TBLs and heat transfer 191
7.04.9.3.4 Structure of the flow near the corners 191
7.04.9.4 Howard's Scaling for High-Ra Convection 191
7.04.9.5 Thermal Convection in More Realistic Systems 192
Acknowledgments 193
References 193

7.04.1 Introduction

Geodynamics is the study of how Earth materials deform and flow over long ($\geq 10^2 - 10^3$ years) time scales. It is thus a science with dual citizenship: at once, a central discipline within the Earth sciences and a branch of fluid dynamics more generally.

The basis of fluid dynamics is a set of general conservation laws for mass, momentum, and energy, which are usually formulated as partial differential equations (Batchelor, 1967, ch. 3). However, because these equations apply to all fluid flows, they describe none in particular and must therefore be supplemented by material constitutive relations and initial and/or boundary conditions that are appropriate for a particular phenomenon of interest. The result, often called a model problem, is the ultimate object of study in fluid mechanics.

Once posed, a model problem can be solved in one of three ways. The first is to construct a physical analog in the laboratory and let nature do the solving. The experimental approach has long played a central role in geodynamics and is discussed in **Chapter 7.03** of this treatise. Another approach is to solve the model problem numerically on a computer, using one of the methods discussed in **Chapter 7.05**. The third possibility, the subject of this chapter, is to solve the problem analytically.

Analytical approaches are of course most effective when the model problem at hand is relatively simple, and lack some of the flexibility of the best experimental and numerical methods. However, they compensate for this by providing a degree of understanding and insight that no other method can match. What is more, they also play a critical role in the interpretation of experimental and numerical results. For example, dimensional analysis is required to ensure proper scaling of experimental and numerical results to the Earth; and local scaling analysis of numerical output can reveal underlying laws that are obscured by numerical tables and graphic images. For all these reasons, the central role that analytical methods have always played in geodynamics is unlikely to diminish.

The purpose of this chapter is to survey the principal analytical methods of geodynamics and the major results that have been obtained using them. These methods are remarkably diverse, and require a correspondingly broad and comprehensive treatment. However, it is equally important to highlight the common structures and styles of argumentation that give analytical geodynamics an impressive unity. In this spirit, I begin with a discussion of the art of formulating geophysical model problems, focusing on three paradigmatic phenomena (heat transfer from magma diapirs, gravitational instability of buoyant layers, and plume–lithosphere interaction, PLI) that will subsequently reappear in the course of the chapter treated by different methods. Thus heat transfer from diapirs is treated using dimensional analysis (**Sections 7.04.3.1** and **7.04.3.2**), scaling analysis (**Section 7.04.3.3**), and boundary-layer theory (**Section 7.04.7.2.2**); gravitational instability using scaling analysis (**Section 7.04.4.4**), long-wave analysis (**Section 7.04.3.3**), and linear stability analysis (**Section 7.04.9.1**); and PLI using lubrication theory and scaling analysis (**Sections 7.04.8.1** and **7.04.8.2**). Moreover, the discussions of these and other phenomena are organized as much as possible around three recurrent themes. The first is the importance of scaling arguments (and the scaling laws to which they lead) as tools for understanding physical mechanisms and applying model results to the Earth. Examples of scaling arguments can be found in **Sections 7.04.3.3, 7.04.4.4, 7.04.7.3, 7.04.8.2, 7.04.9.2.5, 7.04.9.2.6, 7.04.9.3.1**, and **7.04.9.3.4**. The second is the ubiquity of self-similar behavior in geophysical flows, which typically occurs in parts of the spatiotemporal flow domain that are sufficiently far from the inhomogeneous initial or boundary conditions that drive the flow to be uninfluenced by the details of those conditions (**Sections 7.04.4.1–7.04.4.3, 7.04.5.3.3, 7.04.5.3.4, 7.04.7.2.1, 7.04.7.4, 7.04.8.1, 7.04.8.2**, and **7.04.9.3.4**). The third theme is asymptotic analysis, in which the smallness of some crucial parameter in the model problem is exploited to simplify the governing equations, often via a reduction of their dimensionality (**Sections 7.04.4.1, 7.04.5.7,**

7.04.7, 7.04.7.3, 7.04.8.1, 7.04.8.3–7.04.8.6, 7.04.9.2.5, 7.04.9.2.6, and 7.04.9.3.2–7.04.9.3.4). While these themes by no means encompass everything the chapter contains, they can serve as threads to guide the reader through what might otherwise appear a trackless labyrinth of miscellaneous methods.

A final aim of the chapter is to introduce some less familiar methods, drawn from other areas of fluid mechanics, that deserve to be better known among geodynamicists. Examples include the use of Papkovich–Fadle eigenfunction expansions (Section 7.04.5.4.1) and complex variables (Section 7.04.5.5) for 2-D Stokes flows, solutions in bispherical coordinates for 3-D Stokes flows (Section 7.04.5.4.3), and multiple-scale analysis of modulated convection rolls (Sections 7.04.9.2.5 and 7.04.9.2.6).

Throughout this chapter, unless otherwise stated, Greek indices range over values 1 and 2; Latin indices range over 1, 2, and 3; and the standard summation convention over repeated subscripts is assumed. Index notation (e.g., u_i, σ_{ij}) and vector/tensor notation ($\mathbf{u}$, $\boldsymbol{\sigma}$) are used interchangeably as convenience dictates. For simplicity, both u_i and $\mathbf{u}$ are called vectors and both σ_{ij} and $\boldsymbol{\sigma}$ tensors, although strictly speaking, u_i and σ_{ij} are the components of $\mathbf{u}$ and $\boldsymbol{\sigma}$, respectively. The notations $(x,y,z)=(x_1,x_2,x_3)$ for Cartesian coordinates and $(u,v,w)=(u_1,u_2,u_3)$ for the corresponding velocity components are equivalent. Unit vectors in given coordinate directions are denoted by symbols $\boldsymbol{e}_x$, $\boldsymbol{e}_r$, etc. Partial derivatives are denoted either by subscripts or by the symbol ∂, and $\partial_i=\partial/\partial x_i$. Thus, for example,

$$T_x=\partial_x T=\partial_1 T=\frac{\partial T}{\partial x}=\frac{\partial T}{\partial x_1}. \qquad [1]$$

The symbol ∇_h is the 2-D gradient operator with respect to the coordinates (x,y) or (in a few cases) the spherical colatitude and longitude (θ,ϕ). The symbols $\Re[\ldots]$ and $\Im[\ldots]$ denote the real and imaginary parts, respectively, of the bracketed quantities. A 'free-slip' surface is one on which both the normal velocity and the shear traction vanish. Frequently used abbreviations are listed in Table 1.

Table 1 Frequently used abbreviations

Abbreviation	*Meaning*
1-D	One-dimensional
2-D	Two-dimensional
3-D	Three-dimensional
BL	Boundary layer
BVP	Boundary-value problem
CMB	Core–mantle boundary
LHS	Left-hand side
MEE	Method of eigenfunction expansions
MMAE	Method of matched asymptotic expansions
ODE	Ordinary differential equation
PDE	Partial differential equation
PLI	Plume–lithosphere interaction
RHS	Right-hand side
R–T	Rayleigh–Taylor
SBT	Slender-body theory
TBL	Thermal boundary layer

7.04.2 Formulating Geodynamical Model Problems: Three Case Studies

Ideally, a geodynamical model should respect two distinct criteria: it should be sufficiently simple that the essential physics it embodies can be easily understood yet sufficiently complex and realistic that it can be used to draw inferences about the Earth. It is seldom easy to satisfy both these desiderata in a single model; and so most geophysicists tend to emphasize one or the other, according to temperament and education.

However, there is a way around this dilemma: to investigate not only a single model, but also a hierarchical series of models of gradually increasing complexity and realism. Such an investigation – whether carried out by one individual or by many – is a cumulative one in which the initial study of highly simplified models provides the physical understanding required to guide the formulation and investigation of more complex models. In many cases, the simpler models in such a series can be solved analytically, whereas the subsequent more realistic models require numerical or experimental methods. To illustrate how the hierarchical approach works in practice, I have chosen three exemplary geophysical phenomena as case studies: heat transfer from mantle diapirs, buoyant instability of thermal boundary layers (TBLs); and the interaction of mantle plumes with the lithosphere. The following discussions emphasize models at the simpler end of the spectrum that are amenable to analytical methods and omit most of the mathematical details to keep the focus on the conceptual structure of the hierarchical approach.

7.04.2.1 Heat Transfer from Mantle Diapirs

Our first example is the ascent of a hot blob or diapir of magma through the lithosphere, a possible mechanism for the formation of island-arc volcanoes (Marsh, 1978; Marsh and Carmichael, 1974). To evaluate this model, one needs to know how far the diapir can move through the colder surrounding material before losing so much of its excess heat that it solidifies. Figure 1 illustrates a series of model problems that can be used to investigate this question.

Probably the simplest model that still retains much of the essential physics (Marsh, 1978) can be formulated by assuming that (1) the diapir is spherical and has a constant radius, (2) the diapir's interior temperature is uniform and (3) does not vary with time, (4) the ascent speed and (5) the temperature of the lithosphere far from the diapir are constants, (6) the lithosphere is a uniform viscous fluid with constant physical properties, and (7) the viscosity of the diapir is much less than that of the lithosphere. The result is the model shown in Figure 1(a), in which an effectively inviscid fluid sphere with radius a and temperature ΔT ascends at constant speed U through a fluid with constant density ρ, thermal diffusivity κ, viscosity ν, and constant temperature $T=0$ far from the sphere. An analytical solution for the rate of heat transfer q from the diapir (Section 7.04.7.2.2.) can now be obtained if one makes the additional (and realistic) assumption that (8) the Péclet number $\mathrm{Pe}\equiv Ua/\kappa\gg 1$, in which case the temperature variations around the leading hemisphere of the diapir are confined

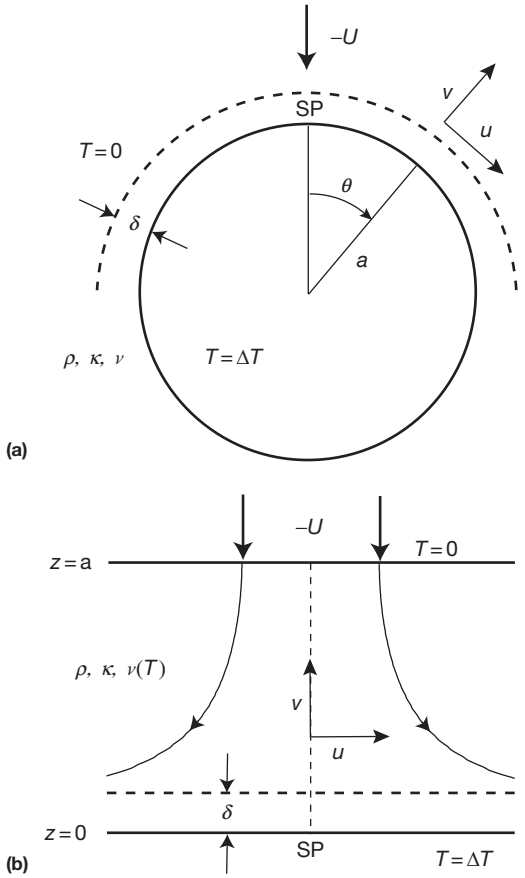


Figure 1 Models for the heat transfer from an ascending magma diapir. (a) Original model in spherical geometry. A sphere of radius a and constant temperature ΔT ascends in an infinite fluid with density ρ, kinematic viscosity ν, thermal diffusivity κ, and temperature $T=0$ far from the diapir. The figure is drawn in the reference frame of the diapir, so that the fluid far from it moves downward with a constant speed $-U$. The colatitude measured from the leading stagnation point (SP) is θ, and the components of the velocity in the colatitudinal and radial directions are u and v, respectively. In the limit $Ua/\kappa \gg 1$, temperature variations in the hemisphere $\theta \leq \pi/2$ are confined to a boundary layer of thickness $\delta \ll a$. The viscosity ν may be constant (Marsh, 1978) or temperature-dependent (Morris, 1982). (b) Modified stagnation-point flow model of Morris (1982). The surface of the hot sphere is replaced by the plane $z=0$, and the far-field streaming velocity $-U$ is imposed as a boundary condition at $z=a$.

to a thin boundary layer of thickness $\delta \ll a$ (**Figure 1(a)**). One thereby finds (see **Section 7.04.3.3** for the derivation)

$$q \sim k_c a \Delta T \mathrm{Pe}^{1/2} \qquad [2]$$

where k_c is the thermal conductivity.

While the model just described provides a first estimate of how the heat transfer scales with the ascent speed and the radius and excess temperature of the diapir, it is far too simple for direct application to the Earth. A more realistic model can be obtained by relaxing assumptions (3) and (5), allowing the temperatures of the diapir and the ambient lithosphere to vary with time. If these variations are slow enough, the heat transfer at each instant will be described by a law of the form (2), but with a time-dependent excess temperature $\Delta T(t)$. A model of this type was proposed by Marsh (1978), who obtained a solution in the form of a convolution integral for the evolving temperature of a diapir ascending through a lithosphere with a prescribed far-field temperature $T_{\mathrm{lith}}(t)$.

A different extension of the simple model of **Figure 1(a)**, also suggested by Marsh (1978), begins from the observation that the viscosity of mantle materials decreases strongly with increasing temperature. A hot diapir will therefore be surrounded by a thin halo of softened lithosphere, which will act as a lubricant and increase the diapir's ascent speed. The effectiveness of this mechanism depends on whether the halo is thick enough, and/or has a viscosity low enough, to carry a substantial fraction of the volume flux $\sim \pi a^2 U$ that the sphere must displace in order to move. Formally, this model is obtained by replacing the constant viscosity ν in **Figure 1(a)** by one that depends exponentially on temperature as $\nu = \nu_0 \exp(-T/\Delta T_r)$, where ΔT_r is a rheological temperature scale.

While this new variable-viscosity model is more realistic and dynamically richer than the original model, its spherical geometry makes analytical solution difficult. However, closer examination reveals that the spherical geometry is not in fact essential: all that matters is that the flow outside the softened halo varies over a characteristic length scale a that greatly exceeds the halo thickness. This recognition led Morris (1982) to study a simpler model in which the flow around the sphere is replaced by a stagnation-point flow between two planar boundaries $z=0$ and $z=a$ (**Figure 1(b)**). The model equations now admit 1-D solutions $T=T(z)$ and $v=v(z)$ for the temperature and the vertical velocity, respectively, which can be determined using the method of matched asymptotic expansions (MMAEs; **Section 7.04.7.3**) in the limit of large viscosity contrast $\Delta T/\Delta T_r \gg 1$ (Morris, 1982). Because the general scaling relationships revealed by the 1-D solution apply equally well to the original spherical geometry, they can now be exploited to simplify the governing equations in spherical coordinates, which can then be solved analytically for certain limiting cases (Ansari and Morris, 1985; Morris, 1982).

7.04.2.2 Plume Formation in TBLs

Our second example (**Figure 2**) is the formation of plumes via the gravitational instability of a horizontal TBL. The first step in formulating a model for this process is to choose a simple representation for the relevant physical properties (density, viscosity, and thermal diffusivity) of the fluid. Because these depend on pressure, temperature, and (possibly) chemical composition, they will vary continuously with depth in the TBL and with time (due to thermal diffusion). As a first approximation, however, one can model the depth distribution of the fluid properties as a nondiffusing and spatially discontinuous two-fluid configuration in which a dense layer with constant thickness h_0, density $\rho_0 + \Delta\rho$, and viscosity η_1 overlies a half-space with density ρ_0 and viscosity η_0 (**Figure 2(a)**). The case of a less dense BL beneath a denser fluid is obtained by turning the system upside down and switching the sign of $\Delta\rho$. Because both configurations are gravitationally unstable, any small perturbation of the interface between the two fluids will grow with time via the Rayleigh–Taylor (R–T) instability. The growth rate of an infinitesimal sinusoidal perturbation with arbitrary wavenumber k can be determined analytically (Selig, 1965;

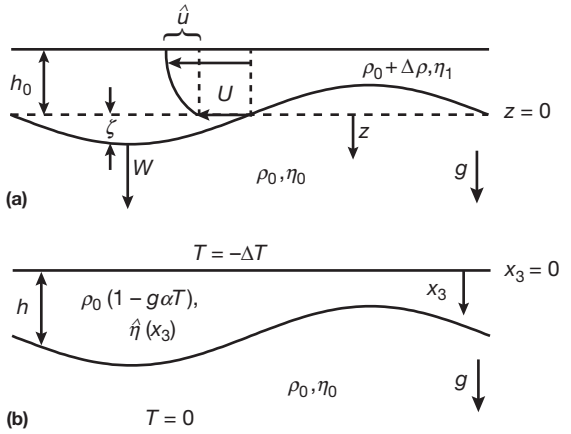


Figure 2 Models for plume formation in a dense/cold thermal boundary layer. (a) Rayleigh–Taylor instability of a layer of fluid with density $\rho_0+\Delta\rho$ and viscosity η_1 above a half-space of fluid with density ρ_0 and viscosity η_0. The initial thickness of the dense layer is h_0, and the deformation of the interface is ζ. The maximum values of the horizontal and vertical velocities at the interface $z=\zeta$ are U and W, respectively, and $\hat{u}$ is the magnitude of the change in horizontal velocity across the layer. (b) Buoyant instability of a cold thermal boundary layer. The upper surface $x_3=0$ of a fluid half-space is held at a fixed temperature $-\Delta T$ relative to the interior. The density of the fluid varies with temperature as $\rho=\rho_0(1-g\alpha T)$, where α is the thermal expansion coefficient. The thickness of the boundary layer is $h(x_1,x_2,t)$, and the viscosity within it is $\hat{\eta}(x_3)$.

Whitehead and Luther, 1975) via a standard linear stability analysis (**Section 7.04.9.1**).

The results described earlier in the text are limited to small disturbances of the interface, suggesting that the next step could be to extend the model to disturbances of finite amplitude. This was done by Canright and Morris (1993), who considered the R–T instability of a layer of fluid with Newtonian or non-Newtonian (power-law) rheology overlying an effectively inviscid half-space. They derived the nonlinear evolution equation governing the thickness of the layer and showed that its solution has a finite-time singularity corresponding to the catastrophic escape of the growing plume. This model is a good example of the trade-off one often faces in developing sophisticated analytical models: an increase in predictive power (e.g., from infinitesimal to finite amplitude) typically comes at the price of a restricted range of validity (e.g., an effectively infinite viscosity contrast between the layer and the half-space).

While simple linearized or finite-amplitude R–T models embody some of the essential physics of plume formation, they are unable to describe such crucial features as the characteristic periodicity of TBL instabilities (Howard, 1964). For this purpose, a more realistic model that incorporates a diffusing temperature field is needed. **Figure 2(b)** shows such a model (Canright, 1987; Lemery et al., 2000), in which a diffusive TBL grows away from a cold surface and subsequently becomes unstable. The density of the fluid varies linearly with temperature, and the viscosity $\hat{\eta}$ can vary as a function of the depth x_3 within the TBL. The additional assumptions that the wavelength of the instability greatly exceeds the thickness of the TBL and that the horizontal components of the fluid velocity are constant across it then permit an analytical reduction of the full 3-D equations to an equivalent set of 2-D equations for the horizontal velocities and the first moment of the temperature in the TBL. The solutions of these equations also exhibit finite-time singularities corresponding to the escape of plumes from the growing boundary layer (**Section 7.04.8.3**).

The foregoing models for plume escape are full-field models that involve the determination of time-dependent 2-D or 3-D velocity fields. A less rigorous but much simpler alternative approach, introduced by Whitehead and Luther (1975), is to idealize the plume as an effectively inviscid sphere whose radius $a(t)$ increases with time as $\dot{a}=Q/(4\pi a^2)$ due to the supply of fluid at a constant volumetric rate Q. The model posits that the sphere detaches from its source when its ascent speed (as predicted by Stokes's law) exceeds $\dot{a}$. This prediction agrees well with laboratory experiments in which the plume fluid is supplied at a constant rate through a pipe. However, it is difficult to apply directly to mantle plumes, which are supplied at a nonconstant rate by a low-viscosity boundary layer. This recognition led Bercovici and Kelly (1997) to generalize Whitehead and Luther's (1975) model to take into account the progressive draining of the boundary layer, yielding a nonlinear evolution equation for the height of the growing plume disturbance.

7.04.2.3 Plume–Lithosphere Interaction

PLI refers to the processes that occur after a rising mantle plume impinges on the base of the lithosphere. Because the plume fluid is buoyant relative to its surroundings, it will spread beneath the lithosphere, eventually forming a thin layer or pancake whose lateral dimensions greatly exceed its thickness.

Figure 3 shows a series of fluid dynamical models that have been used to study PLI, beginning with the kinematical model of Sleep (1987) (**Figure 3(a)**). Sleep's insight was that the flow associated with a plume rising beneath a moving plate can be regarded as the sum of two parts: a (horizontal) radial flow representing buoyant plume fluid emanating from a steady localized source at the top of the plume conduit and an ambient mantle wind in the direction of the plate motion. Fluid from the source can travel only a finite distance upstream against the wind before being blown back downstream again, leading to the formation of a stagnation point (labeled SP in **Figure 3(a)**) at which the wind speed just equals the speed of radial outflow from the source. The stagnation streamline that passes through this point (heavy line in **Figure 3(a)**) divides the (x,y) plane into an inner region containing fluid from the source and an outer region containing fluid brought in from upstream by the wind. The stagnation streamline resembles the shape of the topographic swell around the Hawaiian Island chain (Richards, Hager, and Sleep, 1988).

While the model of Sleep (1987) nicely illustrates the kinematics of plume–ridge interaction (PRI), it neglects the (driving) buoyancy force and (resisting) viscous force that control the spreading of the plume pancake. We now seek the simplest possible model that embodies these dynamics. Following the logic of **Section 7.04.2.2**, we replace the continuous variation of fluid properties by a two-fluid structure, comprising a pancake with constant viscosity η spreading in an ambient fluid

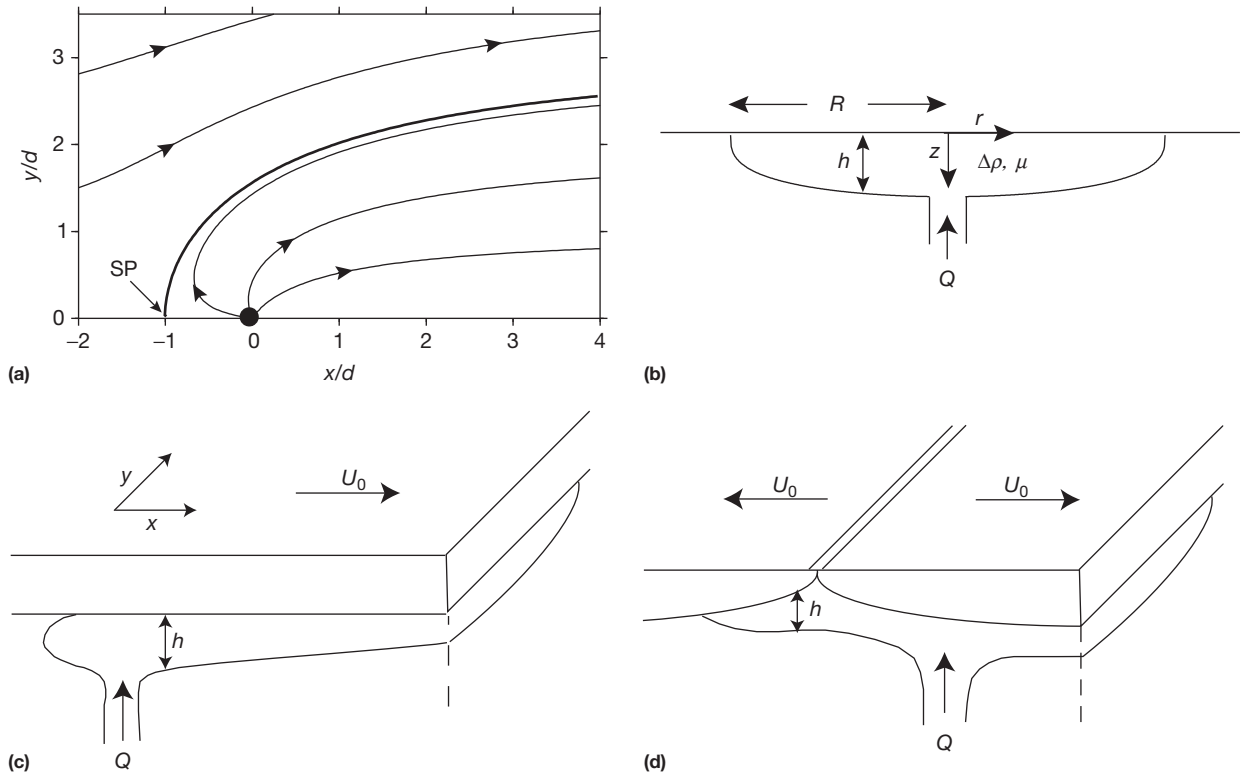


Figure 3 Hierarchy of models for plume–lithosphere interaction. (a) Steady streamlines for the 2-D kinematic model of Sleep (1987). The source is indicated by the black circle, the heavy solid line is the stagnation streamline, and d is the distance between the source and the stagnation point (SP). (b) Spreading beneath a rigid surface of an axisymmetric current of viscous fluid with viscosity η and anomalous density $-\Delta\rho$, supplied at a volumetric rate Q. (c) Same as (b), but beneath a rigid plate moving at speed U_0. (d) Same as (c), but beneath two plates separated by a spreading ridge with half-spreading rate U_0.

with viscosity η_m. We further assume that η_m/η is not too large, so that the influence of the ambient mantle on the spreading of the plume pancake can be neglected. Another simplifying assumption (to be relaxed later) is that the plate does not move with respect to the source of plume fluid. Finally, we suppose that the strength of the source is constant. The result is the viscous gravity current model (Huppert, 1982), wherein fluid with viscosity η and anomalous density $-\Delta\rho$ relative to its surroundings is supplied at a constant volumetric rate Q and spreads beneath a stationary rigid surface (**Figure 3(b)**). The analytical solution for the closely related problem of a current with constant volume V is discussed in **Section 7.04.8.1**.

The next step is to generalize the gravity current model by allowing the plate to move with a constant velocity U relative to the source (Olson, 1990). The resulting refracted plume model (**Figure 3(c)**) is in essence a dynamically self-consistent extension of the 2-D kinematical model of Sleep (1987) and is discussed further in **Section 7.04.8.2**. The refracted plume model can be generalized still further while retaining its analytical character by allowing the plume material to have a more realistic non-Newtonian (power-law) rheology (Asaadi et al., 2011).

As a final illustration, **Figure 3(d)** shows a further extension of the refracted plume model in which the uniform plate is replaced by two plates separated by a spreading ridge. Despite the increased complexity of this PRI model, some analytical results can still be obtained by scaling analysis (Ribe, Christensen and Theissing, 1995).

7.04.3 Dimensional and Scaling Analysis

The goal of studying model problems in fluid mechanics is typically to determine functional relations, or scaling laws, that obtain between certain parameters of interest and the various other parameters in which they depend. Two powerful methods that can be used for this purpose are dimensional analysis and scaling analysis.

7.04.3.1 Buckingham's Π-Theorem and Dynamical Similarity

Dimensional analysis begins from the principle that the validity of physical laws cannot depend on the system of units (SI, cgs, etc.) in which they are expressed. An important consequence of this principle is the Π-theorem of Buckingham (1914).

Suppose that there exists a (generally unknown) functional relationship among N parameters $P_1\ P_2, \ldots, P_N$ that characterize some physical system of interest, namely

$$f_1(P_1, P_2, \ldots, P_N) = 0. \qquad [3]$$

In fluid mechanical problems, the parameters P_n typically include material properties (viscosity, density, etc.), the dimensions of the system (depth, width, etc.), the parameters that drive the flow (the gravitational acceleration, a velocity imposed at a boundary, etc.), spatial coordinates (if one is

concerned with the properties of the flow at a particular point), and the time (if the problem is unsteady). These parameters usually have dimensions but can sometimes be dimensionless (e.g., the angle between the walls of a fluid-filled wedge). Let $M<N$ be the number of the parameters P_n that have independent physical dimensions (physical systems can never have $M=N$, because a dimensionally consistent functional relationship among the parameters P_n is impossible in that case). In most cases, M is just the number of independent primary units that enter into the problem, for example, $M=3$ for mechanical problems involving units of meter, kilogram, and seconds and $M=4$ for thermomechanical problems involving temperature (units of kelvin) in addition. However, there are some exceptions. For example, suppose we wish to determine how the energy E of a body depends on its mass m and on the speed of light c. Naively counting the primary units (m, kg, and s) involved in these quantities, we would conclude that $M=3$, leading to the impossible result $N=M$. However, counting the number of parameters with independent dimensions gives the correct result $M=2$.

Given the numbers N and M defined as above, Buckingham's Π-theorem states that there exist $N-M$ independent dimensionless combinations (often called dimensionless groups) of the original dimensional physical parameters P_n. The theorem gets its name from the tradition of denoting these groups by the Greek letter Π with a subscript $i=1, 2, \ldots, N-M$. The theorem implies that the original dimensional functional relation [3] is entirely equivalent to a dimensionless relation of the form

$$f_2(\Pi_1, \Pi_2, \ldots, \Pi_{N-M}) = 0. \qquad [4]$$

Note that whereas the original function f_1 in eqn [3] had N arguments, the function f_2 in eqn [4] has only $N-M$ arguments. The equivalence of all systems of units requires that each dimensionless group Π_i be a product of powers of the dimensional parameters P_n; no other functional form preserves the value of the dimensionless group when the system of units is changed. The function f_2 in eqn [4], by contrast, can have any form and need not be simple (e.g., a power law). While the total number of independent groups Π_i is fixed ($\equiv N-M$), the definitions of the individual groups are arbitrary and can be chosen as convenient. A more detailed discussion and proof of the Π-theorem can be found in Barenblatt (1996, ch. 1).

The Π-theorem is the basis for the concept of dynamical similarity, according to which two physical systems behave similarly (i.e., proportionally) if they have the same values of the dimensionless groups Π_i that define them. The crucial point is that two systems may have identical values of Π_i even though they are of very different sizes, that is, even if the values of the dimensional parameters P_n are very different. Dynamical similarity is thus a natural generalization of the concept of geometrical similarity, whereby, for example, two triangles of different sizes are similar if they have the same values of the dimensionless parameters (angles and ratios of sides) that define them. Geometrical similarity is a necessary, but not a sufficient, condition for dynamical similarity.

The importance of dynamical similarity for physical modeling is that it allows results obtained in the laboratory or on a computer to be applied to another system with very different scales of length, time, etc. Its power derives from the fact that the function f_2 in eqn [4] has M fewer arguments than the original function f_1. Thus an experimentalist or numerical analyst who seeks to determine how a target dimensional parameter P_1 depends on the other $N-1$ parameters need not vary all of the latter individually; it suffices to vary only $N-M-1$ dimensionless parameters. Consequently, if the variation of a given dimensional parameter requires ≈ 10 samplings, then the use of the Π-theorem reduces the effort involved in searching the parameter space by a factor $\approx 10^M$ (Barenblatt, 1996; see also **Chapter 7.03**). By the same token, the Π-theorem makes possible a far more economical representation of experimental or numerical data. As an example, suppose that we have $N=5$ dimensional parameters, P_1-P_5, from which $N-M=2$ independent dimensionless groups Π_1 and Π_2 can be formed. To represent our data without the help of the Π-theorem, we would need many shelves (one for each value of P_5) each containing many books (one for each value of P_4) each containing many pages (one for each value of P_3) each containing a plot of P_2 versus P_1. By using the Π-theorem, however, we can collapse the whole library onto a single plot of Π_2 versus Π_1.

As a simple illustration of the Π-theorem, consider again the model for heat transfer from a hot sphere (**Figure 1(a)**). Suppose that we wish to determine the radial temperature gradient β (proportional to the local conductive heat flux) as a function of position on the sphere's surface. A list of all the relevant parameters includes the following eight (units in parentheses): β (K m^{-1}), a (m), U (m s^{-1}), ρ (kg m^{-3}), ν (m^2 s^{-1}), κ (m^2 s^{-1}), ΔT (K), and θ (dimensionless). However, ρ can be eliminated immediately because it is the only parameter that involves units of mass: no dimensionless group containing it can be defined. The remaining parameters are $N=7$ in number, $M=3$ of which (e.g., a, ΔT, and U) have independent units, so $N-M=4$ independent dimensionless groups can be formed. It is usually good practice to start by defining a single group containing the target parameter (β here). While inspection usually suffices, one can also proceed more formally by writing the group (say, Π_1) as a product of the desired parameter (β) and unknown powers of any set of M parameters with independent dimensions, for example, $\Pi_1 = \beta a^{n_1} \Delta T^{n_2} U^{n_3}$. The requirement that Π_1 be dimensionless then implies $n_1=1$, $n_2=-1$, and $n_3=0$. Additional groups are then obtained by applying the same procedure to the remaining dimensional parameters in the list (κ and ν in this case). Finally, any remaining parameters in the list that are already dimensionless (θ in this case) can be used as groups by themselves. For the hot sphere, the result is

$$\frac{\beta a}{\Delta T} = \mathrm{fct}\left(\frac{Ua}{\kappa}, \frac{Ua}{\nu}, \theta\right) \qquad [5]$$

where fct is an unknown function. The groups $Ua/\kappa \equiv \mathrm{Pe}$, $Ua/\kappa \equiv \mathrm{Re}$, and $\beta a/\Delta T \equiv \mathcal{N}$ are traditionally called the Péclet number, the Reynolds number, and the (local) Nusselt number, respectively (the last to be distinguished from the global Nusselt number $\mathrm{Nu} \equiv \int_S \mathcal{N}\, \mathrm{d}S$ that measures the total heat flux across the sphere's surface S). As we remarked earlier, the definitions of the dimensionless groups in a relation like eqn [5] are not unique. Thus, one can replace any group by the product of itself and arbitrary powers of the other groups, for example, Pe in eqn [5] by the Prandtl number $\nu/\kappa \equiv \mathrm{Pe}/\mathrm{Re}$.

Furthermore, it often (but not always!) happens that the target parameter ceases to depend on a dimensionless group whose value is very large or very small. For example, in a very viscous fluid such as the mantle, $\mathrm{Re} \ll 1$ because inertia is negligible, and so, Re no longer appears as an argument in eqn [5].

7.04.3.2 Nondimensionalization

When the equations governing the dynamics of the problem at hand are known, another method of dimensional analysis becomes available: nondimensionalization. We illustrate this using the same example of a hot sphere.

The first step is to write down the governing equations, together with all the relevant initial and boundary conditions. Because the problem is both steady and axisymmetric, the dependent variables are the velocity $u(r,\theta)$, the pressure $p(r,\theta)$, and the temperature $T(r,\theta)$, where r and θ are the usual spherical coordinates. If viscous dissipation of energy is negligible, the governing equations and boundary conditions are (see Chapter 7.02)

$$\nabla\cdot\mathbf{u}=0,\quad \mathbf{u}\cdot\nabla T=\kappa\nabla^2 T,\quad \mathbf{u}\cdot\nabla\mathbf{u}=-\rho^{-1}\nabla p+\nu\nabla^2\mathbf{u}, \qquad [6a]$$

$$T(a,\theta)-\Delta T=T(\infty,\theta)=\mathbf{u}(a,\theta)=\mathbf{u}(\infty,\theta)+U\mathbf{e}_z=0, \qquad [6b]$$

where $\mathbf{e}_z$ is a vertical unit vector. The next step is to define dimensionless variables using appropriate problem-specific scales. The best scales to use are not always immediately obvious, and in such cases, a reliable technique is to render all variables dimensionless using unknown scales whose definitions will be specified later. Thus, in the case of the hot sphere, we define dimensionless (primed) variables as

$$\mathbf{x}=L^\star\mathbf{x}',\quad \mathbf{u}=U^\star\mathbf{u}',\quad p=P^\star p',\quad T=T^\star T' \qquad [7]$$

where the scales $L^\star$, $U^\star$, $P^\star$, and $T^\star$ are all unknown. Substituting eqn [7] into eqn [6] and immediately dropping the primes to avoid notational overload, we obtain

$$\frac{U^\star}{L^\star}\nabla\cdot\mathbf{u}=0,\quad \frac{U^\star T^\star}{L^\star}\mathbf{u}\cdot\nabla T=\frac{\kappa T^\star}{L^{\star 2}}\nabla^2 T, \qquad [8a]$$

$$\frac{U^{\star 2}}{L^\star}\mathbf{u}\cdot\nabla\mathbf{u}=-\frac{P^\star}{\rho L^\star}\nabla p+\frac{\nu U^\star}{L^{\star 2}}\nabla^2\mathbf{u}, \qquad [8b]$$

$$T^\star T\left(\frac{a}{L^\star},\theta\right)-\Delta T=T^\star T(\infty,\theta)=0, \qquad [8c]$$

$$U^\star\mathbf{u}\left(\frac{a}{L^\star},\theta\right)=U^\star\mathbf{u}(\infty,\theta)+U\mathbf{e}_z=0. \qquad [8d]$$

The final step is to choose the unknown scales so as to simplify the equations and boundary conditions as much as possible. Inspection of the boundary conditions [8c] and [8d] immediately suggests the choices $L^\star=a$, $U^\star=U$, and $T^\star=\Delta T$. The remaining unknown scale $P^\star$ can now be chosen by setting the coefficient of $-\nabla p$ in eqn [8b] ($=P^\star/\rho L^\star$) equal to either the coefficient of the inertia term $\mathbf{u}\cdot\nabla\mathbf{u}(=U^{\star 2}/L^\star)$ or the coefficient of the viscous term $\nabla^2\mathrm{u}(=\nu U^\star/L^{\star 2})$. The first choice yields $P^\star=\rho U^2$ and makes the dimensionless pressure of order unity in the limit of fast (inertia-dominated) flow. The second choice yields $P^\star=\rho\nu U/a$ and makes the dimensionless pressure of order unity in the limit of slow flow with negligible inertia. With this second choice, we obtain the dimensionless BVP

$$\nabla\cdot\mathbf{u}=0,\quad \mathrm{Pe}\ \mathbf{u}\cdot\nabla T=\nabla^2 T,\quad \mathrm{Re}\ \mathbf{u}\cdot\nabla\mathbf{u}=-\nabla p+\nabla^2\mathbf{u}, \qquad [9a]$$

$$T(1,\theta)-1=T(\infty,\theta)=\mathbf{u}(1,\theta)=\mathbf{u}(\infty,\theta)+\mathbf{e}_z=0 \qquad [9b]$$

where Pe and Re are the Péclet and Reynolds numbers defined in Section 7.04.3.1. If we had made the first choice $P^\star=\rho U^2$, the momentum equation in eqn [9a] would have been

$$\mathbf{u}\cdot\nabla\mathbf{u}=-\nabla p+\mathrm{Re}^{-1}\nabla^2\mathbf{u}. \qquad [10]$$

The dimensionless BVP obtained using $P^\star=\rho U^2$ thus contains exactly the same dimensionless groups (Pe and Re) as the one obtained using $P^\star=\rho\nu U/a$. This is a special case of the more general principle that all possible nondimensionalizations of the equations governing a given physical system yield the same number of independent dimensionless groups, as required by the Π-theorem.

Now, because Pe and Re are the only dimensionless parameters appearing in eqns [9a] and [9b], the dimensionless temperature must have the form $T=T(r,\ \theta,\ \mathrm{Pe},\ \mathrm{Re})$. Differentiating this with respect to r and evaluating the result on the surface $r=1$ to obtain the quantity $\beta a/\Delta T$, we find the same result [5] as we did using the Π-theorem.

Whether one chooses to do dimensional analysis using the Π-theorem or nondimensionalization depends on the problem at hand. The Π-theorem is of course the only choice if the governing equations are not known, but its effective use then requires a good intuition of what the relevant physical parameters are. When the governing equations are known, nondimensionalization is usually the best choice, as the relevant physical parameters appear explicitly in the dimensionless equations and initial/boundary conditions. Further detailed examples of nondimensionalization can be found in Ribe and Davaille (2013).

However, there are cases where it makes sense to use the Π-theorem even when the governing equations are known. This is best explained by way of an example: the slow (inertia-free) rise of a buoyant 'creeping plume' with viscosity η_2 in an ambient fluid with viscosity η_1 (Olson and Singer, 1985). In the laboratory, such plumes can be generated by injecting the plume fluid through a hole of radius a in the bottom of a tank containing the ambient fluid (Figure 4). The fluid issues from the hole with a parabolic (Poiseuille) velocity profile with maximum value U_0 and generates a plume of height $L(t)$, where t is the time since the beginning of the injection.

The equations governing this problem are well known: they are the Stokes (slow viscous flow) equations in each of the two fluids, together with the velocity- and stress-matching conditions on the (unknown and time-dependent) interface between them and the kinematic condition expressing the fact that the interface is a material surface. In addition, the boundary condition on the bottom of the tank is

$$\mathbf{u}^{(1)}=0\quad (r>a), \qquad [11a]$$

$$\mathbf{u}^{(2)}=U_0\left[1-\left(\frac{r}{a}\right)^2\right]\mathbf{e}_z\quad (r\le a) \qquad [11b]$$

where r is the radial distance from the center of the hole and $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$ are the velocities of the ambient fluid and the plume fluid, respectively. Finally, there is an initial condition stating that the interface between the two fluids is flat across the hole at $t=0$.

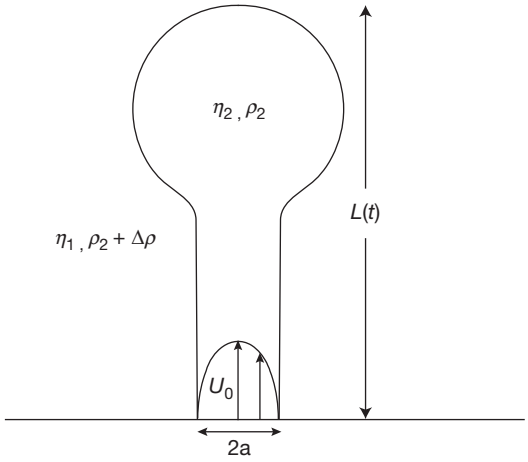


Figure 4 A creeping plume with viscosity η_2 and density ρ_2 ascends in an ambient fluid with viscosity η_1 and density $\rho_2+\Delta\rho$. The plume fluid is injected through a hole of radius a in the bottom of a tank and has a parabolic (Poiseuille) velocity profile with maximum velocity U_0. The time-dependent height of the plume is $L(t)$.

The equations just enumerated can be nondimensionalized using a, U_0, $\eta_1 U_0/a$, and a/U_0 as scales for length, velocity, pressure, and time, respectively. As the saying goes, the details are left as an exercise for the reader. The result is that the dimensionless scaling law for the plume height has the form

$$\frac{L}{a} = \mathrm{fct}\left(\frac{\eta_2}{\eta_1}, \frac{a^2 g\Delta\rho}{\eta_1 U_0}, \frac{U_0 t}{a}\right) \qquad [12]$$

where $\Delta\rho \equiv \rho_1 - \rho_2$ is the density difference between the fluids.

Although the law [12] is perfectly correct, it is potentially misleading because it was derived blindly without applying any physical insight. However, laboratory experiments show that the dynamics and morphology of starting plumes only become interesting after the plume has ascended a distance that greatly exceeds the hole radius a. At such distances, the top of the plume is no longer influenced by the finite radius of the hole, which ceases to be a relevant length scale for the dynamics. The hole then acts as a point source characterized by a single parameter, the volumetric flux $Q \equiv \pi a^2 U_0/2$ that it supplies. The height L of the plume now depends only on the parameters Q, t, η_1, η_2, and $g\Delta\rho$, where the fluid densities appear only in the combination $g\Delta\rho$ because fluids in creeping flow have gravitational mass but effectively zero inertial mass. Applying the Π-theorem to the aforementioned list of six parameters, three of which have independent dimensions, we find

$$\Pi_1 = \mathrm{fct}(\Pi_2, \Pi_3), \qquad [13a]$$

$$\Pi_1 = \left(\frac{\eta_1}{Qg\Delta\rho}\right)^{1/2}\frac{L}{t}, \quad \Pi_2 = \left(\frac{g\Delta\rho}{\eta_1}\right)^{1/2} Q^{1/6} t^{2/3}, \quad \Pi_3 = \frac{\eta_2}{\eta_1}. \qquad [13b]$$

Finally, in the limit $\eta_2/\eta_1 \ll 1$ corresponding to a low-viscosity plume, we anticipate that the plume viscosity η_2 is no longer relevant, in which case eqn [13] reduces to $\Pi_1 = \mathrm{fct}(\Pi_2)$. This last prediction is verified by laboratory measurements in the limit $\eta_2/\eta_1 \ll 1$, which define a universal curve on a plot of Π_1 versus Π_2 (Olson and Singer, 1985, Fig. 6). This simple result would have been much harder to find starting from eqn [12], which can lead the experimenter to suppose incorrectly that a is always a relevant length scale.

The preceding example illustrates nicely both the power and the limitations of dimensional analysis. Its power flows from its capacity to reduce substantially the number of parameters that an experimentalist or numerical analyst must vary in order to understand a physical system. But when performed in a blindly algorithmic way, it can lead the unwary practitioner to retain dynamically irrelevant parameters or (in cases without known governing equations) to overlook parameters that are in fact relevant. The best approach is a flexible and opportunistic one that combines formal dimensional analysis with physical insight in continual dialog with one another.

7.04.3.3 Scaling Analysis

Except when $N - M = 1$, dimensional analysis yields a relation involving an unknown function of one or more dimensionless arguments. To determine the functional dependence itself, methods that go beyond dimensional analysis are required. The most detailed information is provided by a full analytical or numerical solution of the problem, but finding such solutions is rarely easy. Scaling analysis is a powerful intermediate method that provides more information than dimensional analysis while avoiding the labor of a complete solution. It proceeds by estimating the orders of magnitude of the different terms in a set of governing equations, using both known and unknown quantities, and then exploiting the fact that the terms must balance (the definition of an equation!) to determine how the unknown quantities depend on the known.

To illustrate this, we consider once again the problem of determining the local Nusselt number $\mathcal{N}$ for the hot sphere, but now in the specific limit, $\mathrm{Re} \ll 1$ and $\mathrm{Pe} \gg 1$. Recall that Re measures the ratio of advection to diffusion of gradients in velocity ($\equiv$ vorticity), while Pe does the same for gradients in temperature. In the limit $\mathrm{Re} \ll 1$, advection of velocity gradients is negligible relative to diffusion everywhere in the flow field, and $\boldsymbol{u}$ is given by the classic Stokes–Hadamard solution for slow viscous flow around a sphere of another fluid (Section 7.04.5.3.2). When $\mathrm{Pe} \gg 1$, temperature gradients are transported by advection with negligible diffusion everywhere except in a thin TBL of thickness $\delta(\theta) \ll a$ around the leading hemisphere where advection and diffusion are of the same order. Because radial temperature gradients greatly exceed surface-tangential gradients within this layer, the temperature distribution there is described by the simplified boundary-layer forms of the continuity and energy equations (cf. Section 7.04.7)

$$a \sin\theta v_r + (u \sin\theta)_\theta = 0, \qquad [14a]$$

$$a^{-1} u T_\theta + v T_r = \kappa T_{rr} \qquad [14b]$$

where $u(r,\theta)$ and $v(r,\theta)$ are the tangential (θ-) and radial (r-) components of the velocity, respectively, and subscripts indicate partial derivatives. Equations [14a] and [14b] are obtained from eqn [160] by setting $\mathrm{x} = a\theta$ and $r = a\sin\theta$.

We begin by determining the relative magnitudes of the velocity components u and v in the BL. While these can be found directly from the Stokes–Hadamard solution, it is more instructive to do a scaling analysis of the continuity eqn [14a]. Now, $v_r \sim \Delta v/\delta$, where Δv is the change in v across the BL; but because $v(a,\theta)=0$, $\Delta v = v$. Similarly, $u_\theta \sim \Delta u/\Delta\theta$, where Δu is the change in u over an angle $\Delta\theta$ of order unity from the forward stagnation point $\theta=0$ toward the equator $\theta=\pi/2$. But because $u(r,0)=0$, $\Delta u = u$. The continuity equation therefore implies

$$v \sim (\delta/a)u. \qquad [15]$$

We turn now to the LHS of the energy eqn [14b], whose two terms represent advection of temperature gradients in the tangential and radial directions, respectively. Now, the radial temperature gradient $T_r \sim \Delta T/\delta$ greatly exceeds the tangential gradient $a^{-1}T_\theta \sim \Delta T/a$; but this difference is compensated by the smallness of the radial velocity $v \sim (\delta/a)u$, and so the terms representing tangential and radial advection are of the same order. The balance of advection and diffusion in the BL is therefore $a^{-1}uT_\theta \sim \kappa T_{rr}$, which together with $T_{rr} \sim \Delta T/\delta^2$ implies

$$\delta^2 \sim \kappa a/u. \qquad [16]$$

It remains only to determine an expression for u, which depends on the ratio γ of the viscosity of the sphere to that of the surrounding fluid. We consider the limiting cases $\gamma \ll 1$ (a sphere with a free-slip surface) and $\gamma \gg 1$ (an effectively rigid sphere). Because the fluid outside the sphere has constant viscosity, $\mathbf{u}$ varies smoothly over a length scale $\sim a$. Within the TBL, therefore, $\mathbf{u}$ can be approximated by the first term of its Taylor series expansion in the radial distance $r-a \equiv \zeta$ away from the sphere's surface. If the sphere is free-slip, $u_\zeta|_{\zeta=0}=0$, implying that $u \sim U$ is constant across the TBL to lowest order. If however the sphere is rigid, $u|_{\zeta=0}=0$, and $u \sim (\zeta/a)U$. The tangential velocity u at the outer edge $\zeta \sim \delta$ of the TBL is therefore

$$u \sim (\delta/a)^n U \qquad [17]$$

where $n=0$ for a free-slip sphere and $n=1$ for a rigid sphere. Substituting eqn [17] into eqn [16] and noting that $N \sim a/\delta$, we obtain

$$N \sim \mathrm{Pe}^{\frac{1}{n+2}} f_n(\theta) \qquad [18]$$

where $f_n(\theta)$ ($n=1$ or 2) are unknown functions. Thus, when $\mathrm{Pe} \gg 1$, $N \sim \mathrm{Pe}^{1/2}$ if the sphere is free-slip and $N \sim \mathrm{Pe}^{1/3}$ if it is rigid. N is greater in the former case because the tangential velocity u, which carries the heat away from the sphere, is $\sim U$ across the whole TBL.

7.04.4 Self-Similarity and Intermediate Asymptotics

In geophysics and in fluid dynamics more generally, one often encounters functions that exhibit the property of scale invariance or self-similarity. As an illustration, consider a function $f(y,t)$ of two arbitrary variables y and t. The function f is self-similar if it can be written in the form

$$f(y,t) = G(t)F\left(\frac{y}{\delta(t)}\right) \qquad [19]$$

where F, G, and δ are functions of a single argument. The combination $\eta \equiv y/\delta(t)$ is called the similarity variable. Self-similarity simply means that curves of f versus y for different values of t can be obtained from a single universal curve $F(\eta)$ by stretching its abscissa and ordinate by factors $\delta(t)$ and $G(t)$, respectively.

Self-similarity is closely connected with the concept of intermediate asymptotics (Barenblatt, 1996). In many physical situations, one is interested in the behavior of a system at intermediate times, long after it has become insensitive to the details of the initial conditions but long before it reaches a final equilibrium state. The behavior of the system at these intermediate times is often self-similar, as we now illustrate using a simple example of conductive heat transfer (Barenblatt, 1996; **Section 7.04.2.1**).

7.04.4.1 Conductive Heat Transfer

Consider the 1-D conductive heat transfer in a rod $y \in [0, L]$ in which the initial temperature is zero everywhere except in a heated segment of length h centered at $y=y_0$ (**Figure 5**). The width of the heated segment is much smaller than the distance to either end of the rod, and the segment is much closer to the left end than to the right end, that is, $h \ll y_0$ and $y_0 \ll L - y_0$. The ends of the rod are held at zero temperature, and its sides are insulated.

The equation and initial/boundary conditions governing the temperature $T(y,t)$ in the rod are

$$T_t = \kappa T_{yy}, \qquad [20a]$$

$$T(y,0) = T_0(y), \quad T(0,t) = T(L,t) = 0 \qquad [20b]$$

where $T_0(y)$ is the concentrated initial temperature distribution. While it is relatively easy to solve eqns [20a] and [20b] numerically for an arbitrary initial temperature $T_0(y)$, such an approach would not reveal the essential fact that the solution has two distinct intermediate asymptotic, self-similar stages. The first obtains long after the temperature distribution has forgotten the details of the initial distribution $T_0(y)$, but long before it feels the influence of the left boundary condition $T(0,t)=0$, that is, for (roughly) $0.1h^2/\kappa \ll t \ll 0.1y_0^2/\kappa$. In this time range, the rod appears effectively infinite, and the integrated temperature anomaly

$$Q = \int_{-\infty}^{\infty} T(y,t)\,\mathrm{d}y \qquad [21]$$

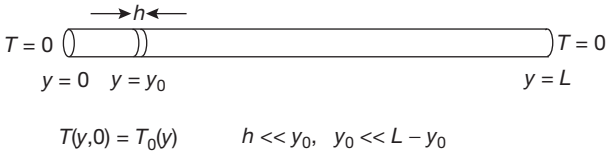


Figure 5 Model for 1-D conductive heat transfer in a rod $y \in [0,L]$ in which the initial temperature is zero everywhere except in a heated segment of length h centered at $y=y_0$. The width of the heated segment is much smaller than the distance to either end of the rod, and the segment is much closer to the left end than to the right end. Both ends of the rod are held at zero temperature, and its sides are insulated.

is constant. The temperature T can depend only on Q, k, t, and $y-y_0$, and only three of these five parameters have independent dimensions. Applying the Π-theorem with $N=5$ and $M=3$, we find

$$T=\frac{Q}{(\kappa t)^{1/2}}F_1(\eta),\quad \eta=\frac{y-y_0}{(\kappa t)^{1/2}} \qquad [22]$$

which is of the general self-similar form [19]. Substituting eqn [22] into eqns [20a] and [21], we find that F_1 satisfies

$$2F_1''+\eta F_1'+F_1=0,\quad \int_{-\infty}^{\infty}F_1\,\mathrm{d}\eta=1. \qquad [22]$$

Upon solving eqn [23] subject to $F_1(\pm\infty)=0$, eqn [22] becomes

$$T=\frac{Q}{2\sqrt{\pi\kappa t}}\exp\left[-\frac{(y-y_0)^2}{4\kappa t}\right]. \qquad [23]$$

The second intermediate asymptotic stage occurs long after the temperature distribution has begun to be influenced by the left boundary condition $T(0,t)=0$, but long before the influence of the right boundary condition $T(L,t)=0$ is felt, or $y_0^2/\kappa \ll t \ll 0.1(L-y_0)^2/\kappa$. During this time interval, the rod is effectively semi-infinite, and the temperature satisfies the boundary conditions

$$T(0,t)=T(\infty,t)=0. \qquad [24]$$

The essential step in determining the similarity solution is to identify a conserved quantity. Multiplying eqn [20a] by y, integrating from $y=0$ to $y=\infty$, and then taking the time derivative outside the integral sign, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^{\infty}yT\,\mathrm{d}y=\kappa\int_0^{\infty}yT_{yy}\,\mathrm{d}y. \qquad [26]$$

However, the RHS of eqn [26] is zero, as can be shown by integrating by parts, applying the conditions [25], and noting that $yT_y|_{y=\infty}=0$ because $T_y\to 0$ more rapidly (typically exponentially) than $y\to\infty$. The temperature moment

$$M=\int_0^{\infty}yT\,\mathrm{d}y \qquad [27]$$

is therefore constant; and because the initial temperature distribution is effectively a delta function concentrated at $y=y_0$, $M=Qy_0$. Now in the time interval in question, the influence of the temperature distribution that existed at the time $\approx 0.1\,y_0^2/\kappa$ when the heated region first reached the near end $y=0$ of the rod will no longer be felt. The temperature will therefore no longer depend on y_0, but only on M, k, y, and $t-t_0$, where t_0 is the effective starting time for the second stage, to be determined later. Applying the Π-theorem as before, we find

$$T=\frac{M}{\kappa(t-t_0)}F_2(\eta),\quad \eta=\frac{y}{\sqrt{\kappa(t-t_0)}}. \qquad [28]$$

Now substitute eqn [28] into eqns [20a] and [27], and solve the resulting equations subject to $F_2(0)=F_2(\infty)=0$, whereupon eqn [28] becomes

$$T=\frac{My}{2\sqrt{\pi}[\kappa(t-t_0)]^{3/2}}\exp\left[-\frac{y^2}{4\kappa(t-t_0)}\right]. \qquad [29]$$

The final step is to determine the starting time $t_0=-y_0^2/6\kappa$ (Barenblatt, 1996, p. 74). Because $t_0<0$, that is, before the rod was originally heated, it represents a virtual starting time with respect to which the behavior of the second stage is self-similar.

7.04.4.2 Self-Similar Buoyant Thermals

Self-similar solutions can also exist for problems involving more than one spatial variable in addition to the time. A geophysically relevant example is Whittaker and Lister's (2008a) solution for a buoyant thermal, a localized but diffusing distribution of heat and buoyancy in an unbounded uniform viscous fluid. If inertia and viscous dissipation of energy are negligible, the governing equations are

$$\nabla\cdot\mathbf{u}=0, \qquad [30a]$$

$$\nu\nabla^2\mathbf{u}=\frac{1}{\rho_0}\nabla p+g[1-\alpha(T-T_0)]\mathbf{e}_z, \qquad [30b]$$

$$\frac{\partial T}{\partial t}+\mathbf{u}\cdot\nabla T=\kappa\nabla^2T \qquad [30c]$$

where a is the thermal expansion coefficient and T_0 and ρ_0 are the temperature and density of the fluid far from the thermal. The total (conserved) buoyancy of the thermal is

$$B=g\alpha\int(T-T_0)\mathrm{d}^3x \qquad [31]$$

where the volume integral is taken over all space. The equations [30a]-[30c] and [31] admit a similarity solution in which lengths scale as $(\kappa t)^{1/2}$, velocities as $(\kappa/t)^{1/2}$, and the temperature anomaly as $B/\alpha g(\kappa t)^{3/2}$. The corresponding similarity transformation is (Whittaker and Lister, 2008a)

$$T(x,t)-T_0=\frac{B}{\alpha g(4\kappa t)^{3/2}}\Theta(\mathbf{X}), \qquad [32a]$$

$$\mathbf{u}(x,t)=\left(\frac{\kappa}{4t}\right)^{1/2}\mathbf{U}(\mathbf{X}), \qquad [32b]$$

$$p(x,t)-p_0+\rho_0gz=\frac{\rho_0\nu}{4t}P(\mathbf{X}), \qquad [32c]$$

where $\mathbf{X}=\mathrm{x}/(4\kappa t)^{1/2}$ is a vector similarity variable for position. Substitution of eqn [32] into eqns [30] and [31] yields

$$\nabla\cdot\mathbf{U}=0, \qquad [33a]$$

$$\nabla^2\mathbf{U}=\nabla P-\mathrm{Ra}\Theta\mathbf{e}_z, \qquad [33b]$$

$$\nabla\cdot[(\mathbf{U}-2\mathbf{X})\Theta]=\kappa\nabla^2\Theta, \qquad [33c]$$

$$\int\Theta\mathrm{d}^3X=1 \qquad [33d]$$

where $\mathrm{Ra}=B/(\kappa\nu)$ is a Rayleigh number and the gradients are taken with respect to $\mathbf{X}$. The appropriate boundary conditions are the vanishing of $\mathbf{U}$, P, and Θ at infinite distance from the thermal. The solution can be obtained analytically if $\mathrm{Ra}\ll 1$, but must be determined numerically otherwise. For $\mathrm{Ra}\ll 1$, the temperature distribution is a spherically symmetrical Gaussian function, while for $\mathrm{Ra}\gg 1$, it comprises a quasispherical head trailed by a wake of length $(\mathrm{Ra}\ln\mathrm{Ra})^{1/2}(\kappa t)^{1/2}$ that contains most of the buoyancy (Whittaker and Lister, 2008a). The latter limit is relevant to the dynamics of starting plumes in the Earth's mantle.

7.04.4.3 Classification of Self-Similar Solutions

Solutions [24] and [29] and the numerical solution of eqn [33] are examples of what Barenblatt (1996) calls self-similar solutions of the first kind, for which dimensional analysis (in some cases supplemented by scaling analysis of the governing equations) suffices to find the similarity variable. They are distinguished from self-similar solutions of the second kind, for which the similarity variable can only be found by solving an eigenvalue problem. We will encounter solutions of the latter type in the sections on viscous corner eddies (Section 7.04.5.3.4) and viscous gravity currents (Section 7.04.8.1).

An example of a self-similar solution that does not fit naturally into either class is the impulsive cooling of a half-space deforming in pure shear. Suppose that the half-space $y \geq 0$ has temperature $T=0$ initially and that, at time $t=0$, the temperature at its surface $y=0$ is suddenly decreased by an amount ΔT. The 2-D velocity field in the half-space is $\mathbf{u} = \dot{\epsilon}(x\mathbf{e}_x - y\mathbf{e}_y)$, where $\dot{\epsilon}$ is the constant rate of extension of the surface $y=0$ and $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors parallel to and normal to the surface, respectively. Given this velocity field, a temperature field $T=T(y,t)$ that is independent of the lateral coordinate x is an allowable solution of the governing advection–diffusion equation $T_t + \mathbf{u}\cdot\nabla T = \kappa\nabla^2 T$, which takes the form

$$T_t - \dot{\epsilon} y T_y = \kappa T_{yy} \qquad [34]$$

subject to the conditions $T(y,0) = T(\infty,t) = T(0,t) + \Delta T = 0$. The limit $\dot{\epsilon} = 0$ corresponds to the classic problem of the impulsive cooling of a static half-space.

Neither dimensional analysis nor scaling analysis is sufficient to determine the similarity variable, which does not have the standard power-law monomial form. However, the solution can be found via a generalized form of the familiar separation-of-variables procedure often used to solve PDEs such as Laplace's equation. Note first that the amplitude ΔT of the temperature in the half-space is a constant. This implies that the function $G(t)$ in the similarity transformation [19] must be independent of time, whence $T(y,t) = \Delta T F(y/\delta(t))$. Substituting this expression into eqn [34] and bringing all terms involving $\delta(t)$ to the LHS, we obtain

$$\frac{\delta\left(\dot{\delta} + \dot{\epsilon}\delta\right)}{\kappa} = -\frac{F''}{\eta F'} \qquad [35]$$

where dots and primes denote differentiation with respect to t and $\eta \equiv y/\delta(t)$, respectively. Now, the LHS of eqn [35] is a function of t only, whereas the RHS depends on y through the similarity variable η. Equation [35] is therefore consistent only if both sides are equal to a constant λ^2, which is positive because $\dot{\delta} > 0$. The solutions for δ and F subject to the conditions $\delta(0) = F(\infty) = F(0) - 1 = 0$ are

$$F = \operatorname{erfc}\frac{\lambda y}{\sqrt{2}\delta}, \quad \delta = \lambda\left\{\frac{\kappa}{\dot{\epsilon}}[1 - \exp(-2\dot{\epsilon}t)]\right\}^{1/2}. \qquad [36]$$

Evidently, λ cancels out when the solution for δ is substituted into the solution for F, because different values of λ merely correspond to different (arbitrary) definitions of the thermal layer thickness δ. With $\dot{\epsilon} = 0$ and the conventional choice $\lambda = \sqrt{2}$, we recover the well-known solution $\delta = 2\sqrt{\kappa t}$ for a static half-space. However, when $\dot{\epsilon} > 0$, the BL thickness approaches a steady-state value $\delta = (2\kappa/\dot{\epsilon})^{1/2}$ for which the downward diffusion of temperature gradients is balanced by an upward advection, and the similarity variable involves an exponential function of time. The only reliable way to find such non-standard self-similar solutions is the separation-of-variables procedure outlined in the preceding text. But the same procedure works just as well for problems with similarity variables of standard form, and will therefore be used throughout this chapter.

In conclusion, we note that similarity transformations can also be powerful tools for reducing and interpreting the output of numerical models. As a simple example, suppose that some such model yields values of a dimensionless parameter W as a function of two dimensionless groups Π_1 and Π_2. Depending on the physics of the problem, it may be possible to express the results in the self-similar form

$$W(\Pi_1, \Pi_2) = F_1(\Pi_1) F_2\left(\frac{\Pi_1}{F_3(\Pi_2)}\right) \qquad [37]$$

where $F_1 - F_3$ are functions to be determined numerically. A representation of the form [37] is not guaranteed to exist, but when it does, it provides a compact way of representing multidimensional numerical data by functions of a single variable that can be fit by simple analytical expressions. An example of the use of this technique for a problem involving five dimensionless groups is the lubrication theory model for PRI of Ribe and Delattre (1998).

7.04.4.4 Intermediate Asymptotics With Respect to Parameters: The R–T Instability

The concept of intermediate asymptotics is not limited to self-similar behavior of systems that evolve in time but also applies in a more general way to functions of one or more parameters that exhibit simple (typically power-law) behavior in some asymptotically defined subregion of the parameter space. Because power-law scaling usually results from a simple dynamical balance between two competing effects, the identification of intermediate asymptotic limits that have this form is crucial for a physical understanding of the system in question.

The dynamical significance of intermediate asymptotic limits and the role that scaling arguments play in identifying them are nicely illustrated by the R–T instability of a fluid layer with density $\rho_0 + \Delta\rho$, viscosity η_0, and thickness h_0 above a fluid halfspace with density ρ_0 and viscosity η_0 (**Figure 2(a)**). The following discussion is adapted from Canright and Morris (1993).

Linear stability analysis of this problem (cf. Section 7.04.9.1) shows that an infinitesimal sinusoidal perturbation $\zeta = \zeta_0 \sin kx$ of the initially flat interface $z=0$ in **Figure 2(a)** grows exponentially at a rate (Whitehead and Luther, 1975)

$$s = s_1 \frac{\gamma}{2\epsilon}\left[\frac{\gamma(C-1) + S - 2\epsilon}{\gamma^2(S+2\epsilon) + 2\gamma C + S - 2\epsilon}\right] \qquad [38]$$

where $s_1 = g\Delta\rho h_0/\eta_1$, $\gamma = \eta_1/\eta_0$, $\epsilon = h_0 k$, $C = \cosh 2\epsilon$ and $S = \sinh 2\epsilon$. Here we shall consider only the long-wavelength limit $\epsilon \ll 1$, for which eqn [38] reduces to

$$\frac{s}{s_1} \sim \frac{\epsilon\gamma(2\epsilon + 3\gamma)}{2(2\epsilon^3 + 3\gamma + 6\epsilon\gamma^2)} \qquad [39]$$

where the viscosity contrast γ is arbitrary.

By noting the ranges of $\gamma(\epsilon)$ for which different pairs of terms in eqn [39] (one each in the numerator and the denominator) are dominant, one finds that eqn [39] has four intermediate asymptotic limits: $\gamma \ll \epsilon^3$, $\epsilon^3 \ll \gamma \ll \epsilon$, $\epsilon \ll \gamma \ll \epsilon^{-1}$, and $\gamma \gg \epsilon^{-1}$. The essential dynamics associated with each are summarized in columns 3–5 of Table 2. Column 3 shows the ratio of the amplitudes of the vertical (W) and horizontal (U) components of the velocity at the interface (Figure 2(a)). As γ increases, the motion of the interface changes from dominantly vertical in limit 1 to dominantly horizontal in limits 3 and 4. The ratio of shear deformation to plug flow in the layer is measured by the ratio $\hat{u}/U$ (column 4), where $\hat{u}$ is the change in horizontal velocity across the layer (Figure 2(a)). The layer deforms mainly by shear in limits 1 and 2 and by plug flow in limits 3 and 4.

The growth rate is determined by whether the interfacial buoyancy $\sim g\Delta\rho\zeta_0$ is supported by the pressure p_1 in the layer or the pressure p_0 in the half-space. Column 5 of Table 2 gives the expression for the pressure that balances the buoyancy in each limit, and column 6 shows the corresponding growth rate $s = W/\zeta_0$. While the pressures can be calculated directly from the analytical solution of the problem (Section 7.04.1), it is more revealing to obtain them via a scaling analysis of the horizontal component $\partial_x p = \eta\nabla^2 u$ of the momentum equation. In the half-space, the only length scale is k^{-1}, so that $\partial_x \sim \partial_z \sim k$. The continuity equation then implies $u \sim w$. The magnitude of $u \sim w$ is set by the larger of the two components of the velocity at the interface, namely, $u \sim [U, W]$, where $[\ldots]$ denotes the maximum of the enclosed quantities. Turning now to the layer, we note that the horizontal and vertical length scales are different, so that $\partial_x \sim k$ and $\partial_z \sim h_0^{-1}$. Moreover, $\nabla^2 u \sim \left[\hat{u}/h_0^2, k^2 U\right]$ is the sum of terms arising from the shear and plug flow components of u. We thereby find

$$p_1 \sim \frac{\eta_1}{h_0^2 k}\left[\hat{u}, \epsilon^2 U\right], \quad p_0 \sim \eta_0 k [U, W]. \qquad [40]$$

In view of the pressure scales [40] and those for W/U and $\hat{u}/U$ from Table 2, the essential dynamics of each of the four intermediate limits can be summarized as follows. In limit 1, the half-space feels the layer as an effectively free (zero shear traction) boundary, the buoyancy is balanced by the pressure $p_0 \sim \eta_0 kW$ in the half-space, and s is controlled by the half-space viscosity η_0. In limit 2, the half-space still sees the layer as free, but the pressure $p_1 \sim \eta_1 \hat{u}/h_0^2 k$ induced by shear flow in the layer is nevertheless sufficient to balance the buoyancy. Because the layer deforms mostly in shear, $\hat{u}$ is related directly to W via the continuity equation ($W \sim \epsilon\hat{u}$), so s is controlled by the layer viscosity η_1. In limit 3, each layer feels the shear stress applied by the other. While the buoyancy is still balanced by the shear-induced pressure in the layer, the dominance of plug flow means that $\hat{u}$ and W are no longer related via the continuity equation, but rather by the matching condition on the shear stress. The growth rate is therefore controlled by the half-space viscosity η_0. Finally, in limit 4, the layer feels the half-space as a free boundary, the buoyancy is balanced by the pressure $p_1 \sim \eta_1 kU$ induced by plug flow in the layer, and s is controlled by the layer viscosity η_1.

The attentive reader will have noticed that the ratios W/U and $\hat{u}/U$ that appear in Table 2 were determined directly from the analytical solution of the problem, and were not themselves found by scaling analysis. The reason is that scaling analysis fails to capture a subtle feature of the shear stress at the interface, which together with the continuity equation determines W/U and $\hat{u}/U$. Consider the interfacial shear stress in fluid 0, which is $\eta_0(\partial_z u_0 + \partial_x w_0)|_{z=0}$. The analytical solution shows that the two terms in this expression scale as

$$\partial_z u_0|_{z=0} \sim \frac{6\epsilon + 2\epsilon^3\gamma}{\chi}, \quad \partial_x w_0|_{z=0} \sim -\frac{3\epsilon^2 + 2\epsilon^3\gamma}{\chi} \qquad [41]$$

where $\chi = 2(6\epsilon + 3\gamma + 2\epsilon^3\gamma^2)$. Summing the two terms in [41] and noting that $\epsilon^2 \ll \epsilon$, we obtain

$$(\partial_z u_0 + \partial_x w_0)|_{z=0} \sim \frac{6\epsilon}{\chi} \qquad [42]$$

The sum [42] is simpler than either of the separate contributions $\partial_z u_0|_{z=0}$ or $\partial_x w_0|_{z=0}$ because the two terms $\sim \pm 2\epsilon^3\gamma/\chi$ in eqn [41] cancel exactly. The scaling of the total tangential stress is therefore significantly different than that of either of its two components $\partial_z u_0|_{z=0}$ or $\partial_x w_0|_{z=0}$. Scaling analysis cannot detect behavior of this kind, because it provides no information on the numerical prefactors of the different terms in a scaling relation. The present example is thus a useful reminder that scaling analysis, while powerful, needs to be used circumspectly and with an awareness of its limitations.

7.04.5 Slow Viscous Flow

Flows with negligible inertia are fundamental in the Earth's mantle, where the Reynolds number $\mathrm{Re} \approx 10^{-20}$. A particularly important subclass of inertialess flow – variously called slow, creeping, or low Reynolds number flow – comprises flows in which the fluid is incompressible, and isothermal and has a rheology with no memory (elasticity). These conditions, while obviously restrictive, are nevertheless sufficiently realistic to have served as a basis for many important geophysical models.

Table 2 The Rayleigh–Taylor instability: Intermediate asymptotic limits

Limit	γ	W/U	$\hat{u}/U$	Balancing pressure	s
1	$\ll \epsilon^3$	ϵ^2/γ	ϵ/γ	$p_0 \sim \eta_0 kW$	$\Delta\rho g/2k\eta_0$
2	$\epsilon^3 \ll \gamma \ll \epsilon$	ϵ^2/γ	ϵ/γ	$p_1 \sim \eta_1 \hat{u}/h_0^2 k \sim \eta_1 W/h_0^3 k^2$	$\Delta\rho g h_0^3 k^2/3\eta_1$
3	$\epsilon \ll \gamma \ll \epsilon^{-1}$	ϵ	ϵ/γ	$p_1 \sim \eta_1 \hat{u}/h_0^2 k \sim \eta_0 W/h_0^2 k$	$\Delta\rho g h_0^2 k/2\eta_0$
4	$\gg \epsilon^{-1}$	ϵ	ϵ^2	$p_1 \sim \eta_1 kU \sim \eta_1 W/h_0$	$\Delta\rho g h_0/4\eta_1$

7.04.5.1 Basic Equations and Theorems

The most general equations required to describe the slow viscous flows discussed in this section are (see Chapter 7.02)

$$\partial_j u_j = 0 \quad [43a]$$

$$\partial_j \sigma_{ij} + b_i = 0 \quad [43b]$$

$$b_i = -\rho \partial_i \chi \quad [43c]$$

$$\partial_{jj}^2 \chi = 4\pi G\rho \quad [43d]$$

$$\sigma_{ij} = -p\delta_{ij} + 2\eta e_{ij}, \quad e_{ij} = \frac{1}{2}\left(\partial_i u_j + \partial_j u_i\right) \quad [43e]$$

$$\eta = \eta_0 (I/I_0)^{1/n-1}, \quad I = \left(e_{ij} e_{ij}\right)^{1/2} \quad [43f]$$

where u_i is the velocity vector, σ_{ij} is the stress tensor, b_i is the gravitational body force per unit volume, χ is the gravitational potential, ρ is the density, p is the pressure, η is the viscosity, and e_{ij} is the strain rate tensor. The quantity I is an invariant of the strain rate tensor that has units of strain rate (s^{-1}) and is directly related to the so-called second or quadratic invariant I_2 ($\equiv -I^2/2$ for a traceless tensor). Equation [43a] is the incompressibility condition. Equation [43b] expresses conservation of momentum in the absence of inertia, and states that the net force (viscous plus gravitational) acting on each fluid element is zero. Equation [43d] is Poisson's equation for the gravitational potential. Equation [43e] is the standard constitutive relation for a viscous fluid. Finally, eqn [43f] is the strain rate-dependent viscosity for a power-law fluid (sometimes called a generalized Newtonian fluid), where η_0 is the viscosity at a reference strain rate $I = I_0$ and n is the power-law exponent. A Newtonian fluid has $n = 1$, while the rheology of dry olivine deforming by dislocation creep is well described by eqn [43f] with $n \approx 3.5$ (Bai et al., 1991). A discussion of more complicated non-Newtonian fluids is beyond the scope of this chapter; the interested reader is referred to Bird et al. (1987).

Viscous flow described by eqn [43] can be driven either externally, by velocities and/or stresses imposed at the boundaries of the flow domain, or internally, by buoyancy forces (internal loads) arising from lateral variations of the density ρ. On the scale of the whole mantle, the influence of long-wavelength lateral variations of ρ on the gravitational potential χ (self-gravitation) is significant and cannot be neglected. In modeling flow on smaller scales, however, one generally ignores Poisson's eqn [43d] and replaces $\nabla\chi$ in eqn [43c] by a constant gravitational acceleration $-\boldsymbol{g}$.

Slow viscous flow exhibits the property of instantaneity: u_i and σ_{ij} at each instant are determined throughout the fluid solely by the distribution of forcing (internal loads and/or boundary motions) acting at that instant. Instantaneity requires that the fluid has no memory (elasticity) and that acceleration and inertia be negligible; there is then no time lag between the forcing and the fluid's response to it. A corollary is that slow viscous flow is quasi-static, any time dependence being due entirely to the time-dependence of the forcing or of the system's geometry.

The theory of slow viscous flow is most highly developed for the special case of Newtonian fluids (Stokes flow), and several excellent monographs on the subject exist (Happel and Brenner, 1991; Kim and Karrila, 1991; Ladyzhenskaya, 1963; Langlois, 1964; Pozrikidis, 1992). Relative to general slow viscous flow, Stokes flow exhibits the important additional properties of linearity and reversibility. Linearity implies that for a given geometry, a sum of different solutions (e.g., for different forcing distributions) is also a solution. It also implies that u_i and σ_{ij} are directly proportional to the forcing that generates them and, hence, for example, that the force acting on a body in Stokes flow is proportional to its speed. Reversibility refers to the fact that changing the sign of the forcing terms reverses the signs of u_i and σ_{ij} for all material particles. The reversibility principle is especially powerful when used in conjunction with symmetry arguments. It implies, for example, that a body with fore-and-aft symmetry falling freely in any orientation in Stokes flow experiences no torque, and that the lateral separation of two spherical diapirs with different radii is the same before and after their interaction (Manga, 1997).

An important theorem concerning Stokes flow is the *Lorentz reciprocal theorem*, which relates two different Stokes flows (u_i, σ_{ij}, b_i) and $(u_i^\star, \sigma_{ij}^\star, b_i^\star)$. This theorem is the starting point for the boundary-integral representation of Stokes flow derived in Section 7.04.5.6.4. Consider the scalar quantity $u_i^\star \partial_j \sigma_{ij}$, which we manipulate as follows:

$$\begin{aligned} u_i^\star \partial_j \sigma_{ij} &= \partial_j\left(u_i^\star \sigma_{ij}\right) - \sigma_{ij}\partial_j u_i^\star = \partial_j\left(u_i^\star \sigma_{ij}\right) - \left(-p\delta_{ij} + 2\eta e_{ij}\right)\partial_j u_i^\star \\ &= \partial_j\left(u_i^\star \sigma_{ij}\right) - 2\eta e_{ij} e_{ij}^\star \end{aligned} \quad [44]$$

By subtracting from eqn [44] analogous expression with the starred and unstarred fields interchanged and setting $\partial_j \sigma_{ij} = -b_i$ and $\partial_j \sigma_{ij}^\star = -b_{ij}^\star$, we obtain the differential form of the Lorentz reciprocal theorem:

$$\partial_j\left(u_i^\star \sigma_{ij} - u_i \sigma_{ij}^\star\right) = u_j b_j^\star - u_j^\star b_j \quad [45]$$

An integral form of the reciprocal theorem is obtained by integrating eqn [45] over a volume V bounded by a surface S and applying the divergence theorem, yielding

$$\int_S u_i^\star \sigma_{ij} n_j \, \mathrm{d}S + \int_V b_j u_j^\star \, \mathrm{d}V = \int_S u_i \sigma_{ij}^\star n_j \, \mathrm{d}S + \int_V b_j^\star u_j \, \mathrm{d}V \quad [46]$$

where n_j is the outward unit normal to S.

Two additional theorems for Stokes flow concern the total rate of energy dissipation

$$E = 2\eta \int_V e_{ij} e_{ij} \mathrm{d}V \quad [47]$$

in a volume V. The first states that the solution of the Stokes equations subject to given boundary conditions is unique, and is most easily proved by showing that the energy dissipated by the difference of two supposedly different solutions is zero (Kim and Karrila, 1991, p. 14). The second is the minimum dissipation theorem, which states that a solution of the Stokes equations for given boundary conditions dissipates less energy than any other solenoidal vector field satisfying the same boundary conditions (Kim and Karrila, 1991, p. 15). Note that this theorem merely compares a Stokes flow with other flows that do not satisfy the Stokes equations. It says nothing about the relative rates of dissipation of Stokes flows with different geometries and/or boundary conditions, and

therefore, its use as a principle of selection among such flows is not justified.

7.04.5.2 Potential Representations for Incompressible Flow

In an incompressible flow satisfying $\nabla\cdot\mathbf{u}=0$, only $N-1$ of the velocity components u_i are independent, where $N=3$ for general 3-D flows and $N=2$ for 2-D and axisymmetric flows. This fact allows one to express all the velocity components in terms of derivatives of $N-1$ independent scalar potentials, thereby reducing the number of independent variables in the governing equations. The most commonly used potentials are the streamfunction ψ (for $N=2$) and the poloidal potential $\mathcal{P}$ and the toroidal potential $\mathcal{T}$ (for $N=3$). Next, we give expressions for the components of $\mathbf{u}$ in terms of these potentials in Cartesian (x,y,z), cylindrical (ρ,ϕ,z), and spherical (r,θ,ϕ) coordinates, together with the PDEs they satisfy for the important special case of constant viscosity.

7.04.5.2.1 2-D and axisymmetric flows

A 2-D flow is one in which the velocity vector $\mathbf{u}$ is everywhere perpendicular to a fixed direction (say, $\mathbf{e}_z$) in space. The velocity can then be represented in terms of a streamfunction by

$$\mathbf{u}=\mathbf{e}_z\times\nabla\psi=-\mathbf{e}_x\psi_y+\mathbf{e}_y\psi_x=\mathbf{e}_\phi\psi_\rho-\frac{\mathbf{e}_\rho}{\rho}\psi_\phi \qquad [48]$$

The PDE satisfied by ψ is obtained by applying the operator $\mathbf{e}_z\times\nabla$ to the momentum eqn [43b] with the constitutive law [43f]. If the viscosity is constant and $b_i=0$, ψ satisfies the biharmonic equation

$$\nabla_h^4\psi=0,\quad \nabla_h^2=\partial_{xx}^2+\partial_{yy}^2=\rho^{-1}\partial_\rho\left(\rho\partial_\rho\right)+\rho^{-2}\partial_{\phi\phi}^2 \qquad [49]$$

An axisymmetric flow (without swirl) is one in which $\mathbf{u}$ at any point lies in the plane containing the point and some fixed axis (say, $\mathbf{e}_z$). In cylindrical and spherical coordinates,

$$\mathbf{u}=\frac{\mathbf{e}_\phi}{\rho}\times\nabla\psi=-\frac{\mathbf{e}_z}{\rho}\psi_\rho+\frac{\mathbf{e}_\rho}{\rho}\psi_z \qquad [50a]$$

$$\mathbf{u}=\frac{\mathbf{e}_\phi}{r\sin\theta}\times\nabla\psi=\frac{\mathbf{e}_\theta}{r\sin\theta}\psi_r-\frac{\mathbf{e}_r}{r^2\sin\theta}\psi_\theta \qquad [50b]$$

where ψ is referred to as the Stokes stream function. The PDE satisfied by the Stokes stream function is obtained by applying the operator $\mathbf{e}_\phi\times\nabla$ to the momentum equation. For a fluid with constant viscosity and no body force, the result is

$$E^4\psi=0,\quad E^2=\frac{h_3}{h_1h_2}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2}{h_1h_3}\frac{\partial}{\partial q_1}\right)+\frac{\partial}{\partial q_2}\left(\frac{h_1}{h_2h_3}\frac{\partial}{\partial q_2}\right)\right] \qquad [51]$$

where (q_1,q_2) are orthogonal coordinates in any half-plane normal to $\mathbf{e}_\phi$, (h_1,h_2) are the corresponding scale factors, and h_3 is the scale factor for the azimuthal coordinate ϕ. Thus $(q_1,q_2,h_1,h_2,h_3)=(z,\rho,1,1,\rho)$ in cylindrical coordinates and $(r,\ \theta,\ 1,\ r,\ r\sin\theta)$ in spherical coordinates. Note that the operator E^2 is in general different from the 2-D Laplacian operator with respect to the same coordinates, namely,

$$\nabla_h^2=\frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2h_3}{h_1}\frac{\partial}{\partial q_1}\right)+\frac{\partial}{\partial q_2}\left(\frac{h_1h_3}{h_2}\frac{\partial}{\partial q_2}\right)\right] \qquad [52]$$

The two operators are identical only for 2-D flows ($h_3=1$).

7.04.5.2.2 3-D flows

The most commonly used potential representation of 3D flows in geophysics is a decomposition of the velocity $\mathbf{u}$ into poloidal and toroidal components (Backus, 1958; Chandrasekhar, 1981). This representation requires the choice of a preferred direction, which is usually taken to be an upward vertical ($\mathbf{e}_z$) or radial ($\mathbf{e}_r$) unit vector. Relative to Cartesian coordinates, the poloidal/toroidal decomposition has the form

$$\begin{aligned}\mathbf{u}&=\nabla\times(\mathbf{e}_z\times\nabla\mathcal{P})+\mathbf{e}_z\times\nabla\mathcal{T}\\&=\mathbf{e}_x\left(-\mathcal{P}_{xz}-\mathcal{T}_y\right)+\mathbf{e}_y\left(-\mathcal{P}_{yz}+\mathcal{T}_x\right)+\mathbf{e}_z\nabla_h^2\mathcal{P}\end{aligned} \qquad [53]$$

where $\mathcal{P}$ is the poloidal potential and $\mathcal{T}$ is the toroidal potential. The associated vorticity $\boldsymbol{\omega}\equiv\nabla\times\mathbf{u}$ is

$$\boldsymbol{\omega}=\mathbf{e}_x\left(\nabla^2\mathcal{P}_y-\mathcal{T}_{xz}\right)+\mathbf{e}_y\left(-\nabla^2\mathcal{P}_x-\mathcal{T}_{yz}\right)+\mathbf{e}_z\nabla_h^2\mathcal{T} \qquad [54]$$

Inspection of eqns [53] and [54] immediately reveals the fundamental distinction between the poloidal and toroidal fields: the former has no vertical vorticity, while the latter has no vertical velocity. Note also that $\mathcal{P}$ and $\mathcal{T}$ are only defined to within arbitrary functions of z, because the velocity associated with such functions is identically zero by eqn [53]. Similarly, an arbitrary harmonic function $f(x,y)$ (i.e., one satisfying $\nabla_h^2 f=0$) can be added to $\mathcal{P}$ without changing the velocity.

The PDEs satisfied by $\mathcal{P}$ and $\mathcal{T}$ are obtained by applying the operators $\nabla\times(\mathbf{e}_z\times\nabla)$ and $\mathbf{e}_z\times\nabla$, respectively, to the momentum eqn [43b] and [43f]. Supposing that the viscosity is constant but retaining a body force $\mathbf{b}\equiv-g\Delta\rho\mathbf{e}_z$ where $\Delta\rho(\boldsymbol{x})$ is a density anomaly, we find

$$\nabla_h^2\nabla^4\mathcal{P}=\frac{g}{\eta}\nabla_h^2(\Delta\rho),\quad \nabla_h^2\nabla^2\mathcal{T}=0. \qquad [55]$$

Note that the equation for $\mathcal{T}$ is homogeneous, implying that flow driven by internal density anomalies in a fluid with constant viscosity is purely poloidal. The same is true in a fluid whose viscosity varies only as a function of depth, although the equations satisfied by $\mathcal{P}$ and $\mathcal{T}$ are then more complicated than eqn [55]. When the viscosity varies laterally, however, the equations for $\mathcal{P}$ and $\mathcal{T}$ are coupled, and internal density anomalies drive a toroidal flow that is slaved to the poloidal flow. Toroidal flow will also be driven by any surface boundary conditions having a nonzero vertical vorticity, even if the viscosity does not vary laterally.

A common problem in geodynamics is to determine a poloidal flow in a fluid layer confined between horizontal rigid and/or free-slip boundaries. The conditions satisfied by $\mathcal{P}$ on these boundaries can be determined as follows. The vanishing of the vertical velocity (impermeability) requires $\nabla_h^2\mathcal{P}=0$, which means that $\mathcal{P}$ must be a harmonic function of x and y on the boundary. But because $\mathcal{P}$ is only defined to within an arbitrary additive harmonic function, we can replace $\nabla_h^2\mathcal{P}=0$ by the simpler condition $\mathcal{P}=0$. Next, if the surface is rigid, then the no-slip condition there requires $\partial_{13}^2\mathcal{P}=\partial_{23}^2\mathcal{P}=0$, which implies that $\partial_3\mathcal{P}$ is constant on the boundary. But because $\mathcal{P}$ is only defined to within an arbitrary additive function of z, we can set the constant equal to zero with no loss of generality to obtain the simpler condition $\partial_3\mathcal{P}=0$. Finally, if the surface is free-slip, then the vanishing of the tangential stress there requires $\partial_1(\nabla_h^2-\partial_{33}^2)\mathcal{P}=\partial_2(\nabla_h^2-\partial_{33}^2)\mathcal{P}=0$. However,

because $\mathcal{P}=0$ on the boundary, these conditions reduce to $\partial^3_{133}\mathcal{P}=\partial^3_{233}\mathcal{P}=0$, which implies that $\partial_{33}\mathcal{P}$ is constant on the boundary. Setting this constant to zero following the same argument as for a rigid surface, we obtain $\partial^2_{33}\mathcal{P}=0$. In summary, the conditions satisfied by $\mathcal{P}$ on horizontal boundaries are

$$\mathcal{P}=\partial^2_{33}\mathcal{P}=0 \quad (\text{free}-\text{slip}) \qquad [56a]$$

$$\mathcal{P}=\partial_3\mathcal{P}=0 \quad (\text{rigid}) \qquad [56b]$$

Because of the Earth's spherical geometry, the spherical-coordinate form of the poloidal/toroidal representation is particularly important in geophysics. However, the definitions of $\mathcal{P}$ and $\mathcal{T}$ used by different authors sometimes differ by a sign and/or a factor of r. Following Forte and Peltier (1987),

$$\begin{aligned}\mathbf{u}&=\nabla\times(r\mathbf{e}_r\times\nabla\mathcal{P})+r\mathbf{e}_r\times\nabla\mathcal{T}\\&=\mathbf{e}_\theta\left[-\frac{1}{r}(r\mathcal{P})_{r\theta}-\frac{\mathcal{T}_\phi}{\sin\theta}\right]+\mathbf{e}_\phi\left[-\frac{1}{r\sin\theta}(r\mathcal{P})_{r\phi}+\mathcal{T}_\theta\right]+\frac{\mathbf{e}_r}{r}\mathcal{B}^2\mathcal{P}\end{aligned} \qquad [57]$$

where

$$\mathcal{B}^2=\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \qquad [58]$$

Another common convention is that of Chandresekhar (1981, Appendix III), who used a poloidal potential $\Phi\equiv-r\mathcal{P}$ and a toroidal potential $\Psi\equiv-r\mathcal{T}$. The PDEs satisfied by $\mathcal{P}$ and $\mathcal{T}$ in a fluid of constant viscosity in spherical coordinates are obtained by applying the operators $\nabla\times(r\mathbf{e}_r\times\nabla)$ and $r\mathbf{e}_r\times\nabla$, respectively, to the momentum eqn [43b], [43f], yielding

$$\mathcal{B}^2\nabla^4\mathcal{P}=\frac{g}{\eta r}\mathcal{B}^2\Delta\rho, \quad \mathcal{B}^2\nabla^2\mathcal{T}=0 \qquad [59]$$

Because the equation [59] for $\mathcal{T}$ is homogeneous, the remarks following eqn [55] apply also to spherical geometry. The boundary conditions [56] also remain valid in spherical geometry if ∂_3 is replaced by ∂_r.

Two additional quantities of interest are the lateral divergence $\nabla_h\cdot\mathbf{u}$ and the radial component $\mathbf{e}_r\cdot(\nabla\times\mathbf{u})$ of the vorticity, which are

$$\nabla_h\cdot\mathbf{u}=-\mathcal{B}^2\left[\frac{1}{r^2}(r\mathcal{P})_r\right], \quad \mathbf{e}_r\cdot(\nabla\times\mathbf{u})=\frac{\mathcal{B}^2\mathcal{T}}{r} \qquad [60]$$

The lateral divergence depends only on the poloidal component of the flow, whereas the radial vorticity depends only on the toroidal component. At the Earth's surface $r=a$, therefore, divergent and convergent plate boundaries (where $\nabla_h\cdot\mathbf{u}\neq0$) are associated with poloidal flow, while transform faults (where $\mathbf{e}_r\cdot(\nabla\times\mathbf{u})\neq0$) reflect toroidal flow. Spherical harmonic decomposition of the Earth's surface velocity field reveals that the power of the toroidal component is 50–90% of the power of the poloidal component at all degrees (O'Connell et al., 1991, Fig. 4). Toroidal motion this intense can only be generated by strong lateral variations in effective viscosity, raising the possibility that the observed poloidal/toroidal partitioning may place constraints on the rheology of the lithosphere. To quantify this link, Bercovici (1993) proposed a semi-kinematic source–sink model of surface flow in which toroidal motion is generated by the interaction of lateral viscosity variations with a prespecified poloidal flow. The solution of the model equations using various candidate non-Newtonian rheologies shows that a power-law rheology is not sufficient to produce plate-like toroidal flow with strong shear localization, and that a more exotic self-lubricating rheology is needed (Bercovici, 1998).

7.04.5.3 Classical Exact Solutions

The equations of slow viscous flow admit exact analytical solutions in a variety of geophysically relevant geometries. Some of the most useful of these solutions are the following:

7.04.5.3.1 Steady unidirectional flow

The simplest conceivable fluid flow is a steady unidirectional flow with a velocity $w(x,y,t)\mathbf{e}_z$, where (x,y) are Cartesian coordinates in the plane normal to $\mathbf{e}_z$. If the viscosity η is constant and inertia is negligible, w satisfies (Batchelor, 1967)

$$\eta\nabla^2_h w=-G \qquad [61]$$

where $-G\equiv p_z$ is a constant pressure gradient. Two special cases are of interest in geophysics. The first is the steady Poiseuille flow through a cylindrical pipe of radius a, for which the velocity W and the volume flux Q are

$$w=\frac{G}{4\eta}(a^2-r^2), \quad Q\equiv2\pi\int_0^a rw\,dr=\frac{\pi Ga^4}{8\eta} \qquad [62]$$

Poiseuille flow in a vertical pipe driven by an effective pressure gradient $g\Delta\rho$ has been widely used as a model for the ascent of buoyant fluid in the conduit or tail of a mantle plume (e.g., Whitehead and Luther, 1975). The second case is that of a 2-D channel $0\le y\le d$ bounded by rigid walls in which flow is driven by a combination of an applied pressure gradient $-G$ and motion of the boundary $y=d$ with speed U_0 in its own plane. For this case,

$$w=\frac{G}{2\eta}y(d-y)+\frac{U_0y}{d}. \qquad [63]$$

An example of a geodynamical application of eqn [63] is the asthenosphere flow model of Yale and Phipps-Morgan (1998).

7.04.5.3.2 Stokes–Hadamard solution for a sphere

Another classical result that is useful in geophysics is the Stokes–Hadamard solution for the creeping flow in and around a fluid sphere with radius a and viscosity η_1 at rest in an unbounded fluid with viscosity $\eta_0=\eta_1/\gamma$ and velocity $-U\mathbf{e}_z$ far from the sphere (Batchelor, 1967, pp. 230–238). In spherical coordinates (r,θ), the outer $(n=0)$ and inner $(n=1)$ stream functions have the forms $\psi_n=-Ua^2\sin^2\theta f_n(r)$, where

$$f_0=\frac{(2+3\gamma)\hat{r}-\gamma\hat{r}^{-1}-2(1+\gamma)\hat{r}^2}{4(1+\gamma)}, \quad f_1=\frac{\hat{r}^2-\hat{r}^4}{4(1+\gamma)} \qquad [64]$$

and $\hat{r}=r/a$. The term $-\hat{r}^2/2$ in the expression for f_0 is the stream function corresponding to the uniform fluid stream with speed $-U$ at $\hat{r}\to\infty$. The drag on the sphere is

$$\mathbf{F}=-4\pi a\eta_2U\frac{1+3\gamma/2}{1+\gamma}\mathbf{e}_z \qquad [65]$$

If the densities of the two fluids differ such that $\rho_1=\rho_0+\Delta\rho$, the steady velocity $\mathbf{V}$ of the sphere as it moves freely under

gravity is obtained by equating **F** to the Archimedean buoyancy force, yielding

$$\mathbf{V} = \frac{a^2 \mathbf{g} \Delta\rho}{3\eta_2} \frac{1+\gamma}{1+3\gamma/2} \qquad [66]$$

where **g** is the gravitational acceleration. The speed of an effectively inviscid sphere ($\gamma = 0$) is only 50% greater than that of a rigid sphere ($\gamma \to \infty$). Moreover, the normal stress jump across the interface is independent of position for any value of γ, implying that the sphere has no tendency to deform even in the complete absence of surface tension (Batchelor, 1967, p. 238). Equation [66] has been widely used to estimate the ascent speed of plume heads (e.g., Whitehead and Luther, 1975) and isolated thermals (e.g., Griffiths, 1986) in the mantle.

7.04.5.3.3 Models for subduction zones and ridges

2-D viscous flow in a fluid wedge driven by motion of the boundaries (corner flow) has been widely used to model mantle flow in subduction zones and beneath mid-ocean ridges. Figure 6(a) and 6(b) show the geometry of the simplest corner-flow models of these features, due, respectively, to McKenzie (1969) and Lachenbruch and Nathenson (1974).

The models of Figure 6(a) and 6(b) admit analytical solutions for both Newtonian and power-law rheologies. The boundary conditions for both models can be satisfied by a self-similar (of the first kind; cf. Section 7.04.4.3) stream function $\psi = -U_0 r F(\phi)$, where (r, ϕ) are polar coordinates. The only nonzero component of the strain rate tensor **e** is $e_{r\phi} = U_0(F + F'')/2r$, whence the invariant $I \equiv (e_{ij}e_{ij})^{1/2}$ in eqn [43f] is simply $\sqrt{2}e_{r\phi}$. The equation satisfied by F is obtained by applying the operator $\mathbf{e}_z \times \nabla$ to the momentum eqn [43b] with the power-law constitutive relation eqn [43e], [43f], where $\mathbf{e}_z$ is a unit vector normal to the flow plane. The result is

$$\left(\frac{d^2}{d\phi^2} + \frac{2n-1}{n^2}\right)\left[\left(\frac{d^2F}{d\phi^2} + F\right)^{1/n}\right] = 0 \qquad [67]$$

where n is the power-law index. The associated pressure, obtained by integrating the $\mathbf{e}_r$-component of the momentum equation with respect to r subject to the condition of vanishing pressure at $r = \infty$, is $p = -\eta U_0 r^{-1}(F' + F''')$, where the viscosity η is given by eqn [43f]. Equation [67] can be solved analytically if n is any positive integer (Tovish et al., 1978). The values most relevant to geophysics (see Chapter 7.02) are $n = 1$ and $n = 3$, for which the general solutions are

$$F_1 = A_1 \sin\phi + B_1 \cos\phi + C_1 \phi \sin\phi + D_1 \phi \cos\phi \qquad [68]$$

$$F_3 = A_3 \sin\phi + B_3 \cos\phi + C_3 H(\phi, D_3) \qquad [69]$$

respectively, where

$$H(\phi, D_3) = 27 \cos\frac{\sqrt{5}}{3}(\phi + D_3) - \cos\sqrt{5}(\phi + D_3) \qquad [70]$$

and $A_n - D_n$ are arbitrary constants for $n = 1$ and $n = 3$ that are determined by the boundary conditions. For the ridge model,

$$\{A_1, B_1, C_1, D_1\} = \frac{\{c^2, 0, 0, -1\}}{\alpha - sc} \qquad [71]$$

$$\{A_3, B_3, C_3, D_3\} = \left\{-C_3(h_1 s + h_1' c),\ \ 0,\ \ \frac{1}{h_1' s - h_1 c},\ \ \frac{3\pi}{2\sqrt{5}}\right\} \qquad [72]$$

where $s = \sin\alpha$, $c = \cos\alpha$, $h_1 = H(\alpha, D_3)$, and $h_1' = H_\phi(\alpha, D_3)$. For the Newtonian ($n = 1$) subduction model in the wedge $0 \le \phi \le \alpha$,

$$\{A_1, B_1, C_1, D_1\} = \frac{\{\alpha s,\ \ 0,\ \ \alpha c - s,\ \ -\alpha s\}}{\alpha^2 - s^2} \qquad [73]$$

For the power-law ($n = 3$) subduction model in $0 \le \phi \le \alpha$, D_3 satisfies $h_1 - h_0 c - h_0' s = 0$ and the other constants are

$$C_3 = \left(h_1' + h_0 s - h_0' c\right)^{-1},\ \ B_3 = -h_0 C_3,\ \ A_3 = -h_0' C_3 \qquad [74]$$

where $h_0 = H(0, D_3)$ and $h_0' = H_\phi(0, D_3)$. If needed, the solutions in the wedge $\alpha \le \phi \le \pi$ can be obtained from eqns [68] and [69] by applying the boundary conditions shown in Figure 6(a). Solution [73] together with the corresponding one for the wedge $\alpha \le \phi \le \pi$ was the basis for Stevenson and Turner's (1977) hypothesis that the angle of subduction is controlled by the balance between the hydrodynamic lifting torque and the opposing gravitational torque acting on the slab. Their results were extended to power-law fluids by Tovish et al. (1978).

An alternative analytical approach for spreading ridges is to use Cartesian coordinates and Fourier transforms in the

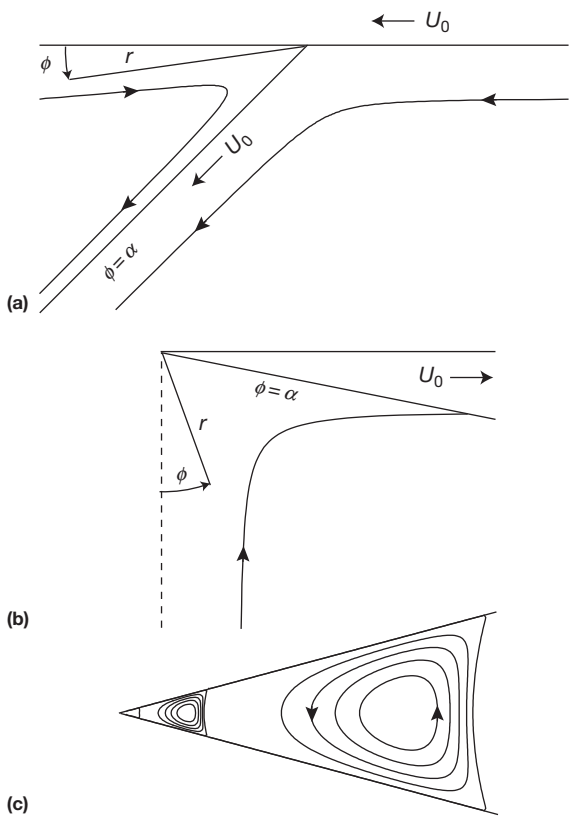


Figure 6 Models for slow viscous flow in wedge-shaped regions. (a) subduction zone; (b) mid-ocean ridge; (c) self-similar viscous corner eddies generated by an agency far from the corner. In all models, streamlines (for Newtonian rheology) are shown by solid lines with arrows.

horizontal directions to calculate the flow induced in the underlying mantle by a specified surface plate velocity field. Ligi et al. (2008) presented solutions of this kind for a spreading ridge with a transform offset above a fluid half-space with constant or layered viscosity.

7.04.5.3.4 Viscous eddies

Another important exact solution for slow viscous flow describes viscous eddies near a sharp corner (Moffatt, 1964). This solution is an example of a self-similar solution of the second kind (Section 7.04.4.3), the determination of which requires the solution of an eigenvalue problem. Here, we consider only Newtonian fluids; for power-law fluids, see Fenner (1975).

The flow domain is a 2-D wedge $|\phi| \leq \alpha$ bounded by rigid walls (Figure 6(c)). Flow in the wedge is driven by an agency (e.g., stirring) acting at a distance $\sim r_0$ from the corner. We seek to determine the asymptotic character of the flow near the corner, that is, in the limit $r/r_0 \to 0$. Because the domain of interest is far from the driving agency, we anticipate that the flow will be self-similar.

The stream function satisfies the biharmonic eqn [49], which admits separable solutions of the form

$$\psi = r^{\lambda} F(\phi) \qquad [75]$$

Substituting eqn [75] into eqn [49], we find that F satisfies

$$F'''' + \left[\lambda^2 + (\lambda-2)^2\right]F'' + \lambda^2(\lambda-2)^2 = 0 \qquad [76]$$

where primes denote d/dϕ. For all λ except 0, 1, and 2, the solution of eqn [76] is

$$F = A\cos\lambda\phi + B\sin\lambda\phi + C\cos(\lambda-2)\phi + D\sin(\lambda-2)\phi \qquad [77]$$

where A–D are arbitrary constants. The solutions for $\lambda = 0$, 1, and 2 do not exhibit eddies (the solution with $\lambda = 1$ is the one used in the models of subduction zones and ridges in Section 7.04.5.3.3).

The most interesting solutions are those for which $\psi(r, \phi)$ is an even function of ϕ. Application of the rigid-surface boundary conditions $F(\pm\alpha) = F'(\pm\alpha) = 0$ to eqn [77] with $B = D = 0$ yields two equations for A and C that have a nontrivial solution only if

$$\sin 2(\lambda-1)\alpha + (\lambda-1)\sin 2\alpha = 0 \qquad [78]$$

The only physically relevant roots of eqn [78] are those with $\Re(\lambda) > 0$, corresponding to solutions that are finite at $r = 0$. When $2\alpha < 146^\circ$, these roots are all complex; let $\lambda_1 \equiv p_1 + iq_1$ be the root with the smallest real part, corresponding to the solution that decays least rapidly toward the corner. Figure 6(c) shows the streamlines $\Re(r^{\lambda}F) =$ constant for $2\alpha = 30^\circ$, for which $p_1 = 4.22$ and $q_1 = 2.20$. The flow comprises an infinite sequence of self-similar eddies with alternating senses of rotation, whose successive intensities decrease by a factor $\exp(\pi p_1/q_1) \approx 414$ toward the corner. In the limit $\alpha = 0$ corresponding to flow between parallel planes, $p_1 = 4.21$ and $q_1 = 2.26$, all the eddies have the same size (≈ 2.78 times the channel width), and the intensity ratio ≈ 348.

In a geophysical context, corner eddies are significant primarily as a simple model for the tendency of forced viscous flows in domains with large aspect ratio to break up into separate cells. An example (Section 7.04.9.3) is steady 2-D cellular convection at high Rayleigh number, in which Stokes flow in the isothermal core of a cell is driven by the shear stresses applied to it by the thermal plumes at its ends. When the aspect ratio (width/depth) $\beta = 1$, the core flow comprises a single cell; but when $\beta = 2.5$, the flow separates into two distinct eddies (Jimenez and Zufiria, 1987, Fig. 2). Busse et al. (2006) and Morris (2008) showed that the cell width can increase significantly if a thin low-viscosity layer is present beneath the upper boundary.

7.04.5.4 Superposition and Eigenfunction Expansion Methods

The linearity of the equations governing Stokes flow is the basis of two powerful methods for solving Stokes flow problems in regular domains: superposition and eigenfunction expansion. In both methods, a complicated flow is represented by infinite sums of elementary separable solutions of the Stokes equations for the coordinate system in question, and the unknown coefficients in the expansion are determined to satisfy the boundary conditions. In the superposition method, the individual separable solutions do not themselves satisfy all the boundary conditions in any of the coordinate directions. In the method of eigenfunction expansion (henceforth MEE), by contrast, the separable solutions are true eigenfunctions that satisfy all the (homogeneous) boundary conditions at both ends of an interval in one of the coordinate directions, so that the unknown constants are determined entirely by the boundary conditions in the other direction(s). Let us turn now to some concrete illustrations of these methods in three coordinate systems of geophysical interest: 2D Cartesian, spherical, and bispherical.

7.04.5.4.1 2D flow in Cartesian coordinates

The stream function ψ for 2-D Stokes flow satisfies the biharmonic eqn [49], which has the general solution

$$\psi = f_1(x+iy) + f_2(x-iy) + (y+ix)f_3(x+iy) + (y-ix)f_4(x-iy) \qquad [79]$$

where $f_1 - f_4$ are arbitrary functions of their (complex) arguments. However, the most useful solutions for applications are the separable solutions that vary periodically in one direction. Setting $\psi \propto \exp ikx$ and solving eqn [49] by separation of variables, we find

$$\psi = [(A_1 + A_2 y)\exp(-ky) + (A_3 + A_4 y)\exp ky]\exp ikx \qquad [80]$$

where $A_1 - A_4$ are arbitrary functions of k. An analogous solution is obtained by interchanging x and y in eqn [80], and both solutions remain valid if k is complex.

Solution [80] has been widely used in geodynamics to describe flows generated by loads that vary sinusoidally in the horizontal direction (e.g., Fleitout and Froidevaux, 1983). The representation of more general flows, however, typically requires the use of a superposition or an eigenfunction expansion. To illustrate the use of these two methods, we consider the problem of driven cavity flow in a rectangular domain $x \in [-\beta/2, \beta/2]$, $y \in [-1/2, 1/2]$ with impermeable ($\psi = 0$) boundaries (Figure 7). The sidewalls $x = \pm\beta/2$ and the bottom $y = -1/2$ are rigid and motionless, while the

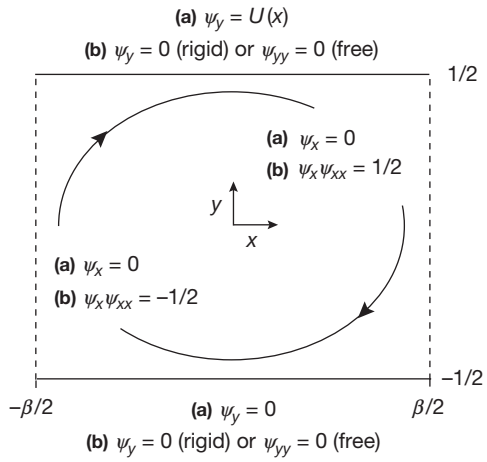


Figure 7 Geometry and boundary conditions for (a) driven cavity flow and (b) steady cellular convection in a rectangle $x \in [-\beta/2, \beta/2]$, $y \in [-1/2, 1/2]$ with impermeable boundaries. All lengths are nondimensionalized by the height of the rectangle.

upper boundary $y = 1/2$ moves in its own plane with velocity $U(x)$ (**Figure 7**, boundary conditions a).

In the superposition method, the solution is represented as a sum of two ordinary Fourier series in the two coordinate directions. The flow is the sum of two parts that are even and odd functions of y, respectively, and the representation for the even part is (Meleshko, 1996)

$$\psi = \sum_{m=1}^{\infty} A_m F(y, p_m, 1)\cos p_m x + \sum_{n=1}^{\infty} B_n F(x, q_n, \beta)\cos q_n y \quad [81a]$$

$$p_m = \frac{2m-1}{\beta}\pi, \quad q_n = (2n-1)\pi \quad [81b]$$

$$F(z, k, r) = \frac{r\tanh(rk/2)\cosh kz - 2z\sinh kz}{2\cosh(rk/2)} \quad [81c]$$

where A_m and B_n are undetermined coefficients. Due to the choice of the wavenumbers p_m and q_n and the function $F(z, k, r)$ (which vanishes at $z = \pm r/2$), both of the series in eqn [81a] satisfy the impermeability condition $\psi = 0$ on all boundaries. Moreover, because the trigonometric systems $\cos p_m x$ and $\cos q_n y$ are complete, the superposition [81] (with suitable choices of A_m and B_n) can represent arbitrary distributions of tangential velocity on $y = \pm 1/2$ and $x = \pm\beta/2$.

Unlike the superposition method, the MEE makes use of the so-called Papkovich–Fadle eigenfunctions $\phi(x)$ that satisfy all the homogeneous boundary conditions in the direction (x in this case) perpendicular to two motionless walls. For simplicity, suppose that $U(x) = U(-x)$, so that ψ is an even function of x. The even eigenfunctions on the canonical unit interval $x \in [-1/2, 1/2]$ are obtained by substituting $\psi = \phi(x)\exp(\lambda y)$ into the biharmonic eqn [49] and solving the resulting equation for ϕ subject to the conditions $\phi(\pm 1/2) = \phi_x(\pm 1/2) = 0$, yielding

$$\phi_n(x) = x\sin\lambda_n x - \frac{1}{2}\tan\frac{\lambda_n}{2}\cos\lambda_n x \quad [82]$$

where λ_n are the first-quadrant complex roots of $\sin\lambda_n + \lambda_n = 0$. The stream function for the flow in the cavity with $x \in [-\beta/2, \beta/2]$ (**Figure 7**) can then be written as (Shankar, 1993)

$$\psi = \sum_{n=1}^{\infty}\left\{A_n\Phi_n\exp\left[-\lambda_n\left(y+\frac{1}{2}\right)\right] + \overline{A}_n\overline{\Phi}_n\exp\left[-\overline{\lambda}_n\left(y+\frac{1}{2}\right)\right] + B_n\Phi_n\exp\left[-\lambda_n\left(\frac{1}{2}-y\right)\right] + \overline{B}_n\overline{\Phi}_n\exp\left[-\overline{\lambda}_n\left(\frac{1}{2}-y\right)\right]\right\} \quad [83]$$

where $\Phi_n = \phi_n(x/\beta)$, overbars denote complex conjugation, and the constants A_n and B_n are chosen to satisfy the boundary conditions at $y = \pm 1/2$. The difficulty, however, is that the reduced biharmonic equation $(d^2/dx^2 + \lambda^2)^2\phi = 0$ satisfied by the eigenfunctions ϕ_n is not self-adjoint, as can be seen by rewriting it in the form

$$\frac{d^2}{dx^2}\begin{pmatrix}\phi\\ -\lambda^{-2}\phi''\end{pmatrix} = \lambda^2\begin{pmatrix}0 & -1\\ 1 & -2\end{pmatrix}\begin{pmatrix}\phi\\ -\lambda^{-2}\phi''\end{pmatrix} \quad [84]$$

and noting that the square matrix on the RHS is not Hermitian. Consequently, the eigenfunctions ϕ_n are not mutually orthogonal, which makes the determination of A_n and B_n a nontrivial matter. One solution is to use a complementary set of adjoint eigenfunctions χ_m which are biorthogonal to the set ϕ_n, although considerable care must then be taken to ensure convergence of the expansion (see Katopodes et al. (2000) for a discussion and references to the relevant literature). A cruder but very effective approach is to determine A_n and B_n via a numerical least-squares procedure that minimizes the misfit of solution [2] to the boundary conditions (Bloor and Wilson, 2006; Shankar, 1993). Shankar (2005) showed how this approach can be extended to an irregular domain by embedding the latter in a larger, regular domain in which a complete set of eigenfunctions exists.

Another geophysically relevant 2-D Stokes problem that can be solved using superposition and eigenfunction expansion methods is that of the flow within the isothermal core of a vigorous (high Rayleigh number) convection cell. The boundary conditions for this case (**Figure 7**, conditions b) comprise either rigid ($\psi_y = 0$) or free-slip ($\psi_{yy} = 0$) conditions at $y = \pm 1/2$ and nonlinear sidewall conditions $\psi_x\psi_{xx} = \pm 1/2$ (derived in **Section 7.04.9.3.2**) that represent the shear stresses applied to the core by the buoyant thermal plumes. If the boundaries $y = \pm 1/2$ are free-slip, the conditions $\psi = \psi_{yy} = 0$ involve only even derivatives of ψ and are satisfied identically if $\psi \propto \cos q_n y$ where q_n is defined by eqn [81b]. The stream function that also satisfies the sidewall impermeability conditions $\psi(\pm\beta/2, y) = 0$ is (Roberts, 1979)

$$\psi = \sum_{n=0}^{\infty} A_n F(x, q_n, \beta)\cos q_n y \quad [85]$$

where $F(x, q_n, \beta)$ is defined by eqn [81c]. The constants A_n are then determined iteratively to satisfy the boundary conditions $\psi_x\psi_{xx}(\pm\beta/2, y) = \pm 1/2$. Strictly speaking, eqn [85] is an eigenfunction expansion, because the functions $\cos q_n y$ satisfy all the homogeneous boundary conditions at $y = \pm 1/2$. In practice, however, the term eigenfunction expansion is usually reserved for expressions like eqn [83] that involve a sequence of complex wave numbers λ_n. No such terminological ambiguity applies for a convection cell bounded by rigid surfaces, for which the flow can be represented by the eigenfunction expansion (Roberts, 1979)

$$\psi = \sum_{n=1}^{\infty}\left[A_n\phi_n(y)\cosh\lambda_n x + \overline{A}_n\overline{\phi}_n(y)\cosh\overline{\lambda}_n x\right] \quad [86]$$

Other examples of the use of superposition and eigenfunction expansion methods for cellular convection problems can be found in Turcotte (1967), Turcotte and Oxburgh (1967), Olson and Corcos (1980), Morris and Canright (1984), Busse et al. (2006), and Morris (2008).

7.04.5.4.2 Spherical coordinates

Lamb (1932) derived a general solution of the equations of Stokes flow in spherical coordinates (r, θ, ϕ). Because the pressure p satisfies Laplace's equation, it may be expressed as a sum of solid spherical harmonics p_l:

$$p = \sum_{l=-\infty}^{\infty} p_l, \quad p_l = r^l \sum_{m=-l}^{l} c_{lm} Y_l^m(\theta, \phi) \qquad [87]$$

where Y_l^m are surface spherical harmonics and c_{lm} are complex coefficients. The velocity $\mathbf{u}$ can then be written as

$$\mathbf{u} = \sum_{l=-\infty}^{\infty} [\nabla \Phi_l + \nabla \times (\mathbf{x}\chi_l)] + \sum_{\substack{l=-\infty \\ l \neq 1}}^{\infty} \left[\frac{(1/2)(l+3) r^2 \nabla p_l - l\mathbf{x} p_l}{\eta (l+1)(2l+3)} \right] \qquad [88]$$

where η is the viscosity, $\mathbf{x}$ is the position vector, and Φ_l and χ_l are solid spherical harmonics of the form [87] but with different coefficients. The first sum in eqn [88] is the solution of the homogeneous Stokes equations $\nabla^2 \mathbf{u} = 0$ and $\nabla \cdot \mathbf{u} = 0$, whereas the second sum is the particular solution of $\nabla^2 \mathbf{u} = \nabla p / \eta$. A recent application of eqn [88] is the solution of Gomilko et al. (2003) for the steady Stokes flow driven by the motion of one of three mutually perpendicular walls that meet in a corner, a 3-D generalization of the 2-D corner-flow model (**Figure 6(a)**). The main difficulty in using Lamb's solution is in applying boundary conditions, because the elements of eqn [88], while complete, do not form an orthogonal basis for vector functions on the surface of a sphere in the way that standard spherical harmonics form an orthonormal basis for scalar functions. Methods for dealing with this problem are discussed in ch. 4 of Kim and Karrila (1991).

7.04.5.4.3 Bispherical coordinates

Flow in a domain bounded by nonconcentric spheres is a useful model for the interaction of one buoyant diapir or plume head with another or with a flat interface (i.e., a sphere of infinite radius). Such flow problems can often be solved analytically using bispherical coordinates (ξ, θ, ϕ), which are related to the Cartesian coordinates (x, y, z) by

$$(x, y, z) = \frac{(a \sin\theta \cos\phi, a \sin\theta \sin\phi, a \sinh\xi)}{\cosh\xi - \cos\theta} \qquad [89]$$

where a is a fixed length scale. Surfaces of constant ξ are nonconcentric spheres with their centers on the axis $x = y = 0$, and $\xi = 0$ corresponds to the plane $z = 0$ (**Figure 8**). For axisymmetric (i.e., independent of ϕ) Stokes flow, the general solution for the stream function ψ and the pressure p is (Stimson and Jeffreys, 1926)

$$\psi = (\cosh\xi - \eta)^{-3/2} \sum_{n=0}^{\infty} C_{n+1}^{-1/2}(\eta) [a_n \cosh\beta_{-1}\xi + b_n \sinh\beta_{-1}\xi + c_n \cosh\beta_3\xi + d_n \sinh\beta_3\xi] \qquad [90a]$$

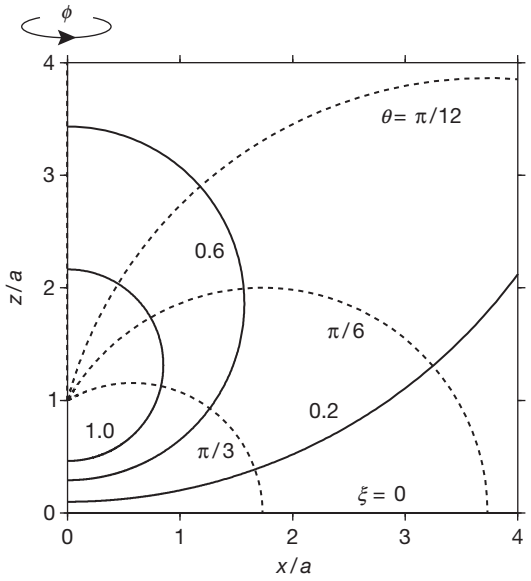


Figure 8 Bispherical coordinates (ξ, θ, ϕ) defined by eqn [89]. Selected lines of constant ξ (solid) and constant θ (dashed) are shown.

$$p = (\cosh\xi - \eta)^{1/2} \sum_{n=0}^{\infty} [A_n \exp\beta_1\xi + B_n \exp(-\beta_1\xi)] P_n(\eta) \qquad [90b]$$

where $\eta = \cos\theta$, $\beta_m = n + m/2$, $C_{n+1}^{-1/2}$ is a Gegenbauer polynomial of order $n+1$ and degree $-1/2$, P_n is a Legendre polynomial, and $a_n - d_n$, A_n, and B_n are arbitrary constants. Koch and Ribe (1989) used a solution of the form [90] to model the effect of a viscosity contrast on the topography and gravity anomalies produced by the motion of a buoyant and deformable fluid sphere beneath a free surface of a fluid with a different viscosity. A representation of general nonaxisymmetric flows in bispherical coordinates was derived by Lee and Leal (1980), and used to determine the flow due to the arbitrary motion of a sphere near a plane wall.

7.04.5.5 The Complex-Variable Method for 2-D Flows

A powerful method for 2-D Stokes flows and analogous problems in elasticity (Muskhelishvili, 1953) is based on the Goursat representation of a biharmonic function ψ in terms of two analytic functions ϕ and χ of the complex variable $z \equiv x + iy$:

$$\psi = \Re[\bar{z}\phi(z) + \chi(z)] \qquad [91]$$

where an overbar denotes complex conjugation. If ψ is the stream function of a 2-D flow, then the velocity components $u = -\psi_y$ and $v = \psi_x$ are

$$v - iu = \phi(z) + z\overline{\phi'(z)} + \overline{\chi'(z)} \qquad [92]$$

the vorticity $\omega \equiv \nabla^2 \psi$ and the pressure p are

$$\omega + ip/\eta = 4\phi'(z) \qquad [93]$$

where η is the viscosity, and the components of the stress tensor are

$$\sigma_{xx} = -2\eta \Im\left[2\phi'(z) - \bar{z}\phi''(z) - \chi''(z)\right] \qquad [94a]$$

$$\sigma_{yy} = -2\eta\Im\left[2\phi'(z) + \bar{z}\phi''(z) + \chi''(z)\right] \quad [94b]$$

$$\sigma_{xy} = 2\eta\Re\left[\bar{z}\phi''(z) + \chi''(z)\right] \quad [94c]$$

The Goursat representation reduces the task of solving the biharmonic equation to one of finding two analytic functions that satisfy the relevant boundary conditions. The method is most powerful when used in conjunction with conformal mapping, which allows a flow domain with a complex shape to be mapped onto a simpler one (such as the interior of the unit circle). A remarkable example is Jeong and Moffat's (1992) analytical solution for the formation of a cusp above a vortex dipole located at depth d beneath the free surface of a viscous fluid. In the geodynamically relevant limit of zero surface tension, the surface displacement $\gamma(x)$ satisfies $x^2\gamma = -(\gamma + 2d/3)^3$, which has an infinitely sharp cusp at $(x, \gamma) = (0, -2d/3)$. Although the model is too idealized for direct application to the mantle, its dynamics are relevant to the formation of cusp-like features by entrainment in thermochemical convection (Davaille, 1999).

7.04.5.6 Singular Solutions and the Boundary Integral Representation

The Stokes equations admit a variety of singular solutions in which the velocity and/or the pressure becomes infinite at one or more points in space. Such solutions are the basis of the boundary-integral representation, whereby a Stokes flow in a given domain is expressed in terms of surface integrals of velocities and stresses over the domain boundaries. The dimensionality of the problem is thereby reduced by one (from 3-D to 2-D or from 2-D to 1-D), making possible a powerful numerical technique – the boundary element method – that does not require discretization of the whole flow domain (Pozrikidis, 1992).

The most useful singular solutions fall into two classes: those involving point forces and those associated with volume sources and sinks. The following discussion is limited to fluids with constant viscosity, but some of the results can be generalized to fluids whose viscosity varies exponentially in one direction (e.g., Zhong, 1996).

7.04.5.6.1 Flow due to point forces

The most important singular solution of the Stokes equations is that due to a point force F_i, or Stokeslet, applied at a position $\mathbf{x}'$ in the fluid. The velocity u_i and stress tensor σ_{ij} induced at any point $\mathbf{x}$ satisfy

$$\partial_j\sigma_{ij} = -F_i\delta(\mathbf{x} - \mathbf{x}'), \quad \partial_j u_j = 0 \quad [95]$$

where $\delta(\mathbf{x} - \mathbf{x}') = \delta(x_1 - x_1')\delta(x_2 - x_2')\delta(x_3 - x_3')$ and δ is the Dirac delta function. Here and throughout Section 7.04.5.6, vector arguments of functions are denoted using boldface vector notation, while other quantities are written using Cartesian tensor (subscript) notation. In an infinite fluid, the required boundary conditions are that $u_i \to 0$ and $\sigma_{ij} \to -p_0\delta_{ij}$ as $|\mathbf{x} - \mathbf{x}'| \to \infty$, where p_0 is the (dynamically irrelevant) far-field pressure. Because the response of the fluid is proportional to the applied force, $u_i = J_{ij}F_j/\eta$ and $\sigma_{ik} = K_{ijk}F_j$, where η is the viscosity and J_{ij} and K_{ijk} are tensorial Green's functions representing the response to a unit force. Substituting these expressions into eqn [95] and eliminating the arbitrary vector F_j, we obtain

$$\partial_i K_{ijk} = -\delta_{jk}\delta(\mathbf{x} - \mathbf{x}'), \quad \partial_i J_{ij} = 0 \quad [96]$$

The solutions of eqn [96] can be found using a Fourier transform (Kim and Karrila, 1991, p. 33) or by reducing eqn [96] to Poisson's equation (Pozrikidis, 1992, p. 22) and are

$$J_{ij}(\mathbf{r}) = \frac{1}{8\pi}\left(\frac{\delta_{ij}}{r} + \frac{r_i r_j}{r^3}\right), \quad K_{ijk}(\mathbf{r}) = -\frac{3}{4\pi}\frac{r_i r_j r_k}{r^5} \quad [97]$$

where $\mathbf{r} = \mathbf{x} - \mathbf{x}'$ and $r = |\mathbf{r}|$. The tensor J_{ij} is often called the Oseen tensor.

The Green's functions [97] represent the 3-D flow generated by a force $\mathbf{F}$ applied at a single point in space. The analogous situation in 2-D is the flow generated by a line force perpendicular to the flow plane with strength $\mathbf{F}$ per unit length. The 2-D analogues of eqn [97] are

$$J_{ij}(\mathbf{r}) = \frac{1}{4\pi}\left(-\delta_{ij}\ln r + \frac{r_i r_j}{r^2}\right), \quad K_{ijk}(\mathbf{r}) = -\frac{1}{\pi}\frac{r_i r_j r_k}{r^4} \quad [98]$$

where the indices i, j, and k now range over values 1 and 2. Expression [98] for J_{ij} diverges logarithmically as $r \to \infty$, which is related to the fact that a 2-D Stokes flow around an infinitely long cylinder does not exist (Stokes's paradox). Soon we will see how this paradox is resolved by the presence of a boundary.

Starting from the 3-D Stokeslet solution, one can use the principle of superposition to construct an infinite variety of additional singular solutions. An example is the flow due to a force dipole, comprising a point force $\mathbf{F}$ at $\mathbf{x}'$ and an equal and opposite force $-\mathbf{F}$ at $\mathbf{x}' - d\mathbf{n}$, where $\mathbf{n}$ is a unit vector directed from the negative to the positive force. The associated velocity field is $u_i(\mathbf{r}) = [J_{ij}(\mathbf{r}) - J_{ij}(\mathbf{r} + d\mathbf{n})]F_j/\eta$. In the limit $d \to 0$ with dF_j fixed,

$$u_i = dF_j n_k G_{ijk}^{\mathrm{FD}}/\eta \quad [99a]$$

$$G_{ijk}^{\mathrm{FD}}(\mathbf{r}) = -\partial_k J_{ij}(\mathbf{r}) \equiv \frac{1}{8\pi}\left[\frac{\delta_{ij}r_k - \delta_{ik}r_j - \delta_{jk}r_i}{r^3} + \frac{3r_i r_j r_k}{r^5}\right] \quad [99b]$$

where G_{ijk}^{FD} is the force-dipole Green's function. The force-dipole moment dn_kF_j is sometimes decomposed into symmetrical (stresslet) and antisymmetric (rotlet) parts (Kim and Karrila, 1991, § 2.5.)

7.04.5.6.2 Flow due to point sources

The basic singular solution of the second class is that associated with a volume source of strength Q at $\mathbf{x}'$, which generates a spherically symmetrical flow

$$u_i = \frac{Qr_i}{4\pi r^3} \quad [100]$$

The flow due to a source doublet comprising a source and sink with equal strengths Q separated by a vector $d\boldsymbol{n}$ pointing from the sink to the source is

$$u_i = Qdn_j G_{ij}^{\mathrm{SD}}, \quad G_{ij}^{\mathrm{SD}}(\mathbf{r}) = -\frac{1}{4\pi}\partial_j\left(\frac{r_i}{r^3}\right) \equiv \frac{1}{4\pi}\left(\frac{3r_i r_j}{r^5} - \frac{\delta_{ij}}{r^3}\right) \quad [101]$$

7.04.5.6.3 Singular solutions in the presence of a boundary

The flow produced by a point force is modified by the presence of an impermeable wall. Consider a force F_i at a point $\mathbf{x}'$ located a distance d from the wall, and let $\mathbf{n}$ be a unit vector normal to

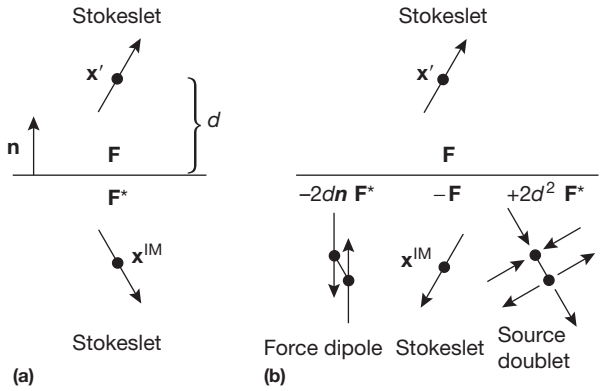


Figure 9 Singular solutions required to describe the flow due to a point force **F** located at a point $\mathbf{x}'$ a distance d above a plane wall. The boundary conditions on the wall (horizontal lines) are satisfied by adding one or more singular solutions at the image point $\mathbf{x}^{\mathrm{IM}} \equiv \mathbf{x}' - 2d\mathbf{n}$, where **n** is the unit vector normal to the wall. The strength or moment of each required singular solution is indicated, and $\mathbf{F}^{\star}$ is the reflection of the vector **F** across the wall. (a) Free-slip wall; (b) rigid wall.

the wall directed toward the side containing $\mathbf{x}'$ (Figure 9(a)). The modified velocity can be written as $u_i = G^{\mathrm{B}}_{ij} F_j/\eta$, where $G^{\mathrm{B}}_{ij}(\mathbf{x}, \mathbf{x}')$ is a Green's function that satisfies all the required boundary conditions at the wall. Its general form is

$$G^{\mathrm{B}}_{ij}(\mathbf{x}, \mathbf{x}') = J_{ij}(\mathbf{x} - \mathbf{x}') + G^{\mathrm{IM}}_{ij}(\mathbf{x} - \mathbf{x}^{\mathrm{IM}}) \qquad [102]$$

where G^{IM}_{ij} is a Green's function that is singular at the image point $\mathbf{x}^{\mathrm{IM}} = \mathbf{x}' - 2d\mathbf{n}$ (Figure 9(a)).

The two limiting cases of greatest interest are free-slip and rigid walls. A free-slip wall is equivalent to a plane of mirror symmetry. The modified flow for this case can therefore be constructed simply by adding a reflected Stokeslet with strength $\mathbf{R}\cdot\mathbf{F} \equiv \mathbf{F}^{\star}$ at the image point $\mathbf{x}^{\mathrm{IM}}$, where $R_{ij} \equiv \delta_{ij} - 2n_i n_j$ is a reflection tensor that reverses the sign of the wall-normal component of a vector while leaving its wall-parallel components unchanged (Figure 9(a)). Therefore,

$$G^{\mathrm{IM}}_{ij}(\mathbf{r}^{\mathrm{IM}}) = R_{jk} J_{ik}(\mathbf{r}^{\mathrm{IM}}) \qquad [103]$$

where $\mathbf{r}^{\mathrm{IM}} = \mathbf{x} - \mathbf{x}^{\mathrm{IM}}$. If the wall is rigid, the no-slip and impermeability conditions can be satisfied by adding three different singular solutions at $\mathbf{x}^{\mathrm{IM}}$: a Stokeslet with strength $-\mathbf{F}$, a force dipole with moment $-2d\mathbf{n}\mathbf{F}^{\star}$, and a source doublet with strength $+2d^2\mathbf{F}^{\star}$ (Figure 9(b)). The result is (Blake, 1971)

$$\begin{aligned} G^{\mathrm{IM}}_{ij}(\mathbf{r}^{\mathrm{IM}}) = &-J_{ij}(\mathbf{r}^{\mathrm{IM}}) - 2dR_{jl}n_k G^{\mathrm{FD}}_{ikl}(\mathbf{r}^{\mathrm{IM}}) \\ &+ 2d^2 R_{jk} G^{\mathrm{SD}}_{ik}(\mathbf{r}^{\mathrm{IM}}) \end{aligned} \qquad [104]$$

where G^{FD}_{ij} and G^{SD}_{ij} are defined by eqns [99b] and [101], respectively.

Expressions analogous to those above also apply to 2-D flow in the presence of a boundary. As an example, consider the case of a free-slip wall $x_2 = 0$ bounding the fluid half-space $x_2 < 0$. Using eqns [102] and [103], we find that the velocity field $u_i = F_j G^{\mathrm{B}}_{ij}/\eta$ is

$$\begin{aligned} u_i = &F_1[J_{i1}(\mathbf{x} - \mathbf{x}') + J_{i1}(\mathbf{x} - \mathbf{x}^{\mathrm{IM}})]/\eta \\ &+ F_2[J_{i2}(\mathbf{x} - \mathbf{x}') - J_{i2}(\mathbf{x} - \mathbf{x}^{\mathrm{IM}})]/\eta \end{aligned} \qquad [105]$$

where $\mathbf{x}^{\mathrm{IM}} = \mathbf{x}' + 2d\mathbf{e}_2$ and J_{ij} is given by eqn [98]. Rewriting eqn [105] in terms of polar coordinates (R, ϕ) and taking the far-field limit $R/d \gg 1$, we obtain

$$\mathbf{u} = \frac{F_1}{2\pi\eta}(\cos\phi\,\mathbf{e}_R - \ln\hat{R}\mathbf{e}_1) + \frac{F_2}{2\pi\eta}\frac{\cos 2\phi}{\hat{R}}\mathbf{e}_R \qquad [106]$$

where $\hat{R} = R/d$ and $\mathbf{e}_R$ is the radial unit vector. Now, if the direction of the line force **F** is perpendicular to the wall ($F_1 = 0$), the velocity [106] vanishes in the limit $\hat{R} \to \infty$. The problem is therefore well posed and Stokes's paradox is resolved. If however **F** has a nonzero wall-parallel component $F_1 \neq 0$, the velocity diverges logarithmically as $\hat{R} \to \infty$.

The difference between the two cases just discussed illustrates a general principle: the solutions of the equations governing 2-D Stokes flow in infinite and semi-infinite domains exist only if the net force on the fluid is zero. This criterion is satisfied for, for example, a rigid circular cylinder moving perpendicular to a free-slip wall because the integrated normal stress over the wall turns out to be exactly equal and opposite to the drag on the cylinder. There is no corresponding solution for a cylinder moving parallel to a free-slip wall because the shear stresses on the wall that would be required to balance the drag are zero by definition. By contrast, a rigid wall can support both normal and shear stresses. The 2-D Stokes equations therefore have solutions for cylinders moving either perpendicular to or parallel to a rigid wall (Wakiya, 1975).

7.04.5.6.4 Boundary-integral representation

The boundary-integral representation for Stokes flow expresses the velocity u_i at any point in a fluid volume V bounded by a surface S in terms of the velocity and traction on S. The starting point is the integral form [46] of the Lorentz reciprocal theorem. Let (u_i, σ_{ij}) be the flow of interest in a fluid with no distributed body forces ($b_i = 0$), and let $u_i^{\star} \equiv J_{ij}(\mathbf{x} - \mathbf{x}')F_j/\eta$ and $\sigma_{ik}^{\star} \equiv K_{ijk}(\mathbf{x} - \mathbf{x}')F_j$ be the flow produced by a point force $b_i^{\star} \equiv F_i\delta(\mathbf{x} - \mathbf{x}')$ at the point $\mathbf{x}'$. Substituting these expressions into eqn [46] and dropping the arbitrary vector F_i, we obtain

$$\begin{aligned} \frac{1}{\eta}\int_S J_{ij}(\mathbf{x} - \mathbf{x}')\sigma_{ik}(\mathbf{x})n_k(\mathbf{x})\mathrm{d}S(\mathbf{x}) - \int_V u_j(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}')\mathrm{d}V(\mathbf{x}) \\ = \int_S K_{ijk}(\mathbf{x} - \mathbf{x}')u_i(\mathbf{x})n_k(\mathbf{x})\mathrm{d}S(\mathbf{x}) \end{aligned} \qquad [107]$$

where the normal **n** points out of V. Now,

$$\int_V u_j(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}')\mathrm{d}V = \chi(\mathbf{x}')u_j(\mathbf{x}') \qquad [108]$$

where $\chi(\mathbf{x}') = 0$, 1/2, or 1 depending on whether $\mathbf{x}'$ lies outside V, right on S, or inside V, respectively. Substituting eqn [108] into eqn [107] and interchanging the roles of $\mathbf{x}$ and $\mathbf{x}'$, we obtain the boundary-integral representation

$$\frac{1}{\eta}\int_{S'} J_{ij}\sigma_{ik}n_k\mathrm{d}S' - \int_{S'} K_{ijk}u_i n_k \mathrm{d}S' = \chi(\mathbf{x})u_j(\mathbf{x}) \qquad [109]$$

where the arguments of the quantities in the integrands ($\mathbf{x}'$ for n_k, σ_{ik}, and u_i; $\mathbf{x}' - \mathbf{x}$ for J_{ij} and K_{ijk}) have been suppressed for brevity and $\mathrm{d}S' = \mathrm{d}S(\mathbf{x}')$. The first integral in eqn [109] represents the velocity due to a surface distribution of point forces with density $\sigma_{ik}n_k dS'$. It is called the single-layer potential by analogy to the electrostatic potential generated by a surface distribution of

electric charges. The second integral, called the double-layer potential, represents the velocity field generated by the sum of a distribution of sources and sinks and a symmetrical distribution of force dipoles (Kim and Karrila, 1991, Section 2.4.2).

An important extension of the integral representation [109] is to the buoyancy-driven motion of a fluid drop with viscosity $\eta_2 \equiv \gamma\eta_1$ and density $\rho_2 \equiv \rho_1 + \Delta\rho$ in another fluid with viscosity η_1 and density ρ_1 (Manga and Stone, 1993; Pozrikidis, 1990). Let S, V_1, and V_2 be the surface of the drop and the volumes outside and inside it, respectively. We begin by writing separate integral equations of the form [109] for each fluid:

$$-\frac{1}{\eta_1}\int_{S'} J_{ij}\sigma_{ik}^{(1)} n_k \mathrm{d}S' + \int_{S'} K_{ijk} u_i^{(1)} n_k \mathrm{d}S' = \chi_1(\mathbf{x}) u_j^{(1)}(\mathbf{x}) \qquad [110a]$$

$$\frac{1}{\eta_2}\int_{S'} J_{ij}\sigma_{ik}^{(2)} n_k \mathrm{d}S' - \int_{S'} K_{ijk} u_i^{(2)} n_k \mathrm{d}S' = \chi_2(\mathbf{x}) u_j^{(2)}(\mathbf{x}) \qquad [110b]$$

where $\mathbf{n}$ points out of the drop and the volume (V_1 or V_2) in which a given quantity is defined is indicated by a superscript in parentheses. Here, $\chi_1(\mathbf{x}) = 0$, 1/2, or 1 if $\mathbf{x}$ is in V_2, right on S, or in V_1, respectively, and $\chi_2(\mathbf{x})$ is defined similarly but with the subscripts 1 and 2 interchanged. Now, multiply eqn [110b] by γ, add the result to eqn [110a], and apply the matching conditions $u_j^{(1)} = u_j^{(2)} = u_j$ and $(\sigma_{ik}^{(1)} - \sigma_{ik}^{(2)}) n_k = n_i \Delta\rho g_k x_k'$ on S, where g_k is the gravitational acceleration. The result is

$$\chi_1(\mathbf{x}) u_j^{(1)}(\mathbf{x}) + \gamma\chi_2(\mathbf{x}) u_j^{(2)}(\mathbf{x}) - (1-\gamma)\int_{S'} K_{ijk} u_i n_k \mathrm{d}S' = -\frac{g_k \Delta\rho}{\eta_1}\int_{S'} J_{ij} n_i x_k' \mathrm{d}S' \qquad [111]$$

For points $\mathbf{x}$ on S, eqn [111] reduces to

$$\frac{1}{2}u_j(\mathbf{x}) - \frac{1-\gamma}{1+\gamma}\int_{S'} K_{ijk} u_i n_k \mathrm{d}S' = -\frac{g_k \Delta\rho}{\eta_1(1+\gamma)}\int_{S'} J_{ij} n_i x_k' \mathrm{d}S' \quad (\mathbf{x} \in S) \qquad [112]$$

a Fredholm integral equation of the second kind for the velocity $\mathbf{u}$ of the interface. Once $\mathbf{u}$ on S has been determined by solving eqn [112], $\mathbf{u}$ at points in V_1 and V_2 can be determined if desired from eqn [111]. A generalization of eqn [112] to $N \geq 2$ interacting drops was derived by Manga and Stone (1993).

The integral eqn [112], which must in general be solved numerically, has been used in geodynamics to model systems comprising distinct fluids with different viscosities. Manga and Stone (1993) solved the N-drop generalization of eqn [112] using the boundary element method (Pozrikidis, 1992) to investigate the buoyancy-driven interaction between two drops in an infinite fluid with a different viscosity. Manga et al. (1993) used a similar method to study the interaction of plume heads with compositional discontinuities in the Earth's mantle. The deformation of viscous blobs in a 2-D cellular flow was investigated by Manga (1996), who concluded that geochemical reservoirs can persist undisturbed for relatively long times if they are 10–100 times more viscous than the surrounding mantle. Manga (1997) showed that the deformation-induced mutual interactions of deformable diapirs in a rising cloud cause the diapirs progressively to cluster but that the rate of this clustering is probably too slow to affect significantly the lateral spacing of rising diapirs in the mantle. Leahy and Bercovici (2010) used the BEM to study the gravitational spreading of a lens of dense low-viscosity melt along the 410 km discontinuity.

A final important application of the BEM is free (buoyancy-driven) subduction of sheets of viscous fluid immersed in another less dense and less viscous fluid. The equation to be solved is still eqn [112], but the interface S is now the surface (both sides plus the ends) of a sheet whose lateral dimensions greatly exceed its thickness. Morra et al. (2007, 2012) developed a multipole-accelerated BEM method for modeling the subduction of several interacting viscous sheets in a spherical annulus beneath a free surface. Morra and Regenauer-Lieb (2006) proposed a hybrid method in which a finite-element solution inside the sheet is coupled to a BEM solution for the outer flow, and Morra et al. (2006) subsequently used it to investigate the causes of the curvature of trenches and island arcs. Ribe (2010) and Li and Ribe (2012) used the BEM to determine scaling laws for the sinking speed and style of free subduction as functions of the sheet/mantle viscosity ratio, the mantle depth/sheet thickness ratio, and the initial slab geometry and interpreted the results using thin viscous-sheet theory (see **Section 7.04.8.4** for further discussion).

7.04.5.7 Slender-Body Theory

Slender-body theory (SBT) is concerned with Stokes flow around thin rodlike bodies whose length greatly exceeds their other two dimensions. The approach takes its departure from Stokes's paradox: the fact that a solution of the equations for Stokes flow around an infinitely long circular cylinder moving steadily in an unbounded viscous fluid does not exist, due to a logarithmic singularity that makes it impossible to satisfy all the boundary conditions (Batchelor, 1967). However, the problem can be regularized in one of three ways: by including inertia, by making the length of the cylinder finite, or by making the domain bounded (**Section 7.04.5.6.3**). SBT is concerned with the second of these possibilities.

The canonical problem of SBT is to determine the force $\mathbf{F}$ on a rod of length 2ℓ and radius $a \ll \ell$ moving with uniform velocity $\mathbf{U}$ in a viscous fluid. The solution can be found using the MMAE, which exploits the fact that the flow field comprises two distinct regions characterized by very different length scales. The first or inner region includes points whose radial distance ρ from the rod is small compared with their distance from the rod's nearer end. In this region, the fluid is not affected by the ends of the rod and sees it as an infinite cylinder with radius a. The second, outer region $\rho \gg a$ is at distances from the rod that are large compared to its radius. Here, the fluid is unaffected by the rod's finite radius and sees it as a line distribution of point forces with effectively zero thickness. The basic idea of the MMAE is to obtain two different asymptotic expansions for the velocity field that are valid in the inner and outer regions, respectively, and then to match them together in an intermediate or overlap region where both expansions must coincide.

As the details of the matching are rather complicated, we defer our discussion of the MMAE to **Section 7.04.7.3**, where it will be applied in the context of boundary-layer theory. Here, we simply quote the lowest-order result that the force on a moving cylinder whose axis is parallel to a unit vector $\mathbf{e}$ is

$$\mathbf{F} = -4\pi\eta\ell\epsilon[2\mathbf{U} - (\mathbf{U}\cdot\mathbf{e})\mathbf{e}] \quad [113]$$

where $\epsilon = [\ln(2\ell/a)]^{-1}$. The force $\mathbf{F}$ includes both drag (parallel to $\mathbf{U}$) and lift (perpendicular to $\mathbf{U}$) components in general. The lift vanishes only when $\mathbf{U}$ is perpendicular to or parallel to $\mathbf{e}$, and the drag is twice as large in the former case as in the latter. For further details and extensions of SBT, see Batchelor (1970), Cox (1970), Keller and Rubinow (1976), and Johnson (1980). Geophysically relevant applications of SBT include Olson and Singer (1985), who used eqn [113] to predict the rise velocity of buoyant quasi-cylindrical (diapiric) plumes. Koch and Koch (1995) used an expression analogous to eqn [113] for an expanding ring to model the buoyant spreading of a viscous drop beneath the free surface of a much more viscous fluid. Whittaker and Lister (2006a) presented a model for a creeping plume above a planar boundary from a point source of buoyancy, in which they modeled the flow outside the plume as that due to a line distribution of Stokeslets. Whittaker and Lister (2008a) used SBT to determine the structure of the self-similar wake trailing a rising buoyant thermal in Stokes flow. Whittaker and Lister (2008b) used SBT to calculate the trajectories of plumes rising and widening diffusively in a background shear flow.

7.04.5.8 Flow Driven by Internal Loads

Because inertia is negligible in the mantle, the flow at each instant is determined entirely by the distribution of internal density anomalies (loads) at that instant. This principle is the basis of a class of internal loading models in which an instantaneous mantle flow field is determined by convolving a load distribution with a Green's function that represents the mantle's response to a unit load. However, it proves convenient here to define the unit load not as a point force, but rather as a surface force concentrated at a single radius and whose amplitude is proportional to a spherical harmonic $Y_l^m(\theta,\phi)$ of degree l and order m. Because the Stokes equations are separable in spherical coordinates, the Green's function representing the response to a harmonic surface load of a given degree and order satisfies an ODE in the radial coordinate r that can be solved analytically. This approach was adumbrated by Pekeris (1935), Runcorn (1967), and McKenzie (1977); formalized by Parsons and Daly (1983) for a plane layer with constant viscosity; and extended by Richards and Hager (1984), Ricard et al. (1984), and Forte and Peltier (1987) to spherical geometry with self-gravitation and radially variable viscosity. To illustrate the method, we now sketch Forte and Peltier's (1987) analytical derivation of the Green's function for a self-gravitating mantle with uniform viscosity.

7.04.5.8.1 Wave-domain Green's functions

Our starting point is eqns [43a]–[43e] that govern slow viscous flow in an incompressible, self-gravitating, Newtonian mantle. Let

$$\rho = \rho_0 + \hat{\rho}(r,\theta,\phi), \quad \boldsymbol{\sigma} = \boldsymbol{\sigma}_0(r) + \hat{\boldsymbol{\sigma}}(r,\theta,\phi), \quad \chi = \chi_0(r) + \hat{\chi}(r,\theta,\phi) \quad [114]$$

where hatted quantities are perturbations of the field variables about a reference hydrostatic state denoted by a subscript 0. Substituting eqn [114] into eqns [43b] and [43d] and neglecting products of perturbation quantities, we obtain

$$0 = \nabla\cdot\hat{\boldsymbol{\sigma}} + \mathbf{g}_0\hat{\rho} - \rho_0\nabla\hat{\chi} \quad [115a]$$

$$\nabla^2\hat{\chi} = 4\pi G\hat{\rho} \quad [115b]$$

where $\mathbf{g}_0 = -\nabla\chi_0$. The second term on the RHS of eqn [115a] represents the buoyancy force acting on the internal density anomalies, and the third term represents the additional force associated with the perturbations in the gravitational potential that they induce (self-gravitation). For consistency with other sections of this chapter, the signs of all gravitational potentials (χ_0, $\hat{\chi}$, etc.) referred to later in the text are opposite to those of Forte and Peltier (1987).

As noted in **Section 7.04.5.2**, the flow driven by internal density anomalies in a fluid with constant or depth-dependent viscosity is purely poloidal. For the constant-viscosity case we are now considering, the poloidal potential $\mathcal{P}(r,\theta,\phi)$ satisfies the first of eqns [59], which remains valid in a self-gravitating mantle. To simplify the notation, we write the poloidal potential and Green's functions without the superposed hat symbol even though they are perturbations relative to the background motionless hydrostatic state. Substitute into the first equation of eqn [59] the expansions

$$\mathcal{P} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\mathcal{P}_l^m(r)Y_l^m(\theta,\phi), \qquad \hat{\rho} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\hat{\rho}_l^m(r)Y_l^m(\theta,\phi) \quad [116]$$

where $Y_l^m(\theta,\phi)$ are surface spherical harmonics satisfying $\mathcal{B}^2Y_l^m = -l(l+1)Y_l^m$ and $\mathcal{B}^2$ is defined by eqn [58]. We thereby find that $\mathcal{P}_l^m(r)$ satisfies

$$D_l^2\mathcal{P}_l^m(r) = \frac{g_0\hat{\rho}_l^m}{\eta r}, \quad D_l = \frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{l(l+1)}{r^2} \quad [117]$$

where the gravitational acceleration g_0 has been assumed constant (Forte and Peltier, 1987). Now, define a poloidal Green's function $P_l(r,r')$ that satisfies

$$D_l^2P_l(r,r') = \delta(r-r') \quad [118]$$

$P_l(r,r')$ represents the poloidal flow generated at the radius r by an infinitely thin density contrast of spherical harmonic degree l and unit amplitude located at radius r'. The poloidal flow due to a distributed density anomaly $\hat{\rho}_l^m(r)$ is then obtained by convolving $\hat{\rho}_l^m(r)$ with the Green's function, yielding

$$\mathcal{P}_l^m(r) = \frac{g_0}{\eta}\int_{a_2}^{a_1}\frac{\hat{\rho}_l^m(r')}{r'}P_l(r,r')\mathrm{d}r' \quad [119]$$

where a_2 and a_1 are the inner and outer radii of the mantle, respectively.

At all radii $r \neq r'$, the Green's function $P_l(r,r')$ satisfies the homogeneous form of eqn [118], which has the general solution

$$P_l(r,r') = A_nr^l + B_nr^{-l-1} + C_nr^{l+2} + D_nr^{-l+1} \quad [120]$$

where $A_n - D_n$ are undetermined constants with $n=1$ for $r' < r \leq a_1$ and $n=2$ for $a_2 \leq r < r'$. These eight constants are determined by the boundary conditions at $r=a_1$ and $r=a_2$ and by matching conditions at $r=r'$. The vanishing of the radial velocity at $r=a_1$ and $r=a_2$ requires

$$P_l(a_1, r') = P_l(a_2, r') = 0 \qquad [121]$$

and the vanishing of the shear stress requires

$$\frac{d^2 P_l}{dr^2}(a_1, r') = \frac{d^2 P_l}{dr^2}(a_2, r') = 0 \qquad [122]$$

Turning now to the matching conditions, we define $[A] \equiv A(r'+) - A(r'-)$ to be the jump in the quantity A across the radius $r = r'$. Continuity of the normal and tangential velocities and the shear stress requires

$$[P_l] = \left[\frac{dP_l}{dr}\right] = \left[\frac{d^2 P_l}{dr^2}\right] = 0 \qquad [123]$$

The normal stress, however, is discontinuous at $r = r'$. By integrating eqn [118] from $r'-$ to $r'+$ and applying eqn [123], we find

$$\left[\frac{d^3 P_l}{dr^3}\right] = 1 \qquad [124]$$

Substitution of eqn [120] into eqns [121]–[124] yields eight equations for $A_n - D_n$, the solutions of which are

$$C_n = -\frac{B_n}{a_n^{2l+3}} = \frac{1}{2(2l+3)(2l+1)(r')^{l-1}} \frac{(a_1/a_n)^{2l+3} - (r'/a_2)^{2l+3}}{1 - (a_1/a_2)^{2l+3}} \qquad [125a]$$

$$D_n = -a_n^{2l-1} A_n = \frac{a_n^{2l-1}}{2(4l^2-1)(r')^{l-3}} \frac{(a_1/a_n)^{2l-1} - (r'/a_2)^{2l-1}}{1 - (a_1/a_2)^{2l-1}} \qquad [125b]$$

The next step is to determine the gravitational potential anomaly. Because the flow induces deformations (dynamic topography) of the Earth's surface and of the core–mantle boundary (CMB), the total potential anomaly is

$$\hat{\chi}_l^m(r) = (\hat{\chi}_0)_l^m(r) + (\hat{\chi}_1)_l^m(r) + (\hat{\chi}_2)_l^m(r) \qquad [126]$$

where $\hat{\chi}_0$, $\hat{\chi}_1$, and $\hat{\chi}_2$ are the potentials associated with the internal load, the deformation $\hat{a}_1$ of the Earth's surface, and the deformation $\hat{a}_2$ of the CMB, respectively. To determine $\hat{\chi}_0$, we note that the general solution of eqn [115b] is

$$\hat{\chi}(\mathbf{x}) = -G \int_V \frac{\hat{\rho}(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} dV' \qquad [127]$$

where $\mathbf{x}$ is the 3-D position vector and the integral is over the whole mantle. We now invoke the expansion (Jackson, 1975, p. 102)

$$|\mathbf{x} - \mathbf{x}'|^{-1} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{1}{2l+1} \frac{r_<^l}{r_>^{l+1}} Y_l^m(\theta, \phi) \overline{Y_l^m}(\theta', \phi') \qquad [128]$$

where $r_< = \min(r, r')$ and $r_> = \max(r, r')$ and an overbar denotes complex conjugation. Substituting eqn [128] into eqn [127] and integrating over θ' and ϕ', we obtain

$$(\hat{\chi}_0)_l^m(r) = -\frac{4\pi G}{2l+1} \int_{a_2}^{a_1} (r')^2 \frac{r_<^l}{r_>^{l+1}} \hat{\rho}_l^m(r') dr' \qquad [129]$$

Expressions for $\hat{\chi}_1$ and $\hat{\chi}_2$ can be obtained from eqn [129] by replacing $\hat{\rho}_l^m(r')$ by $(\rho_0 - \rho_w)(\hat{a}_1)_l^m \delta(r' - a_1)$ and $(\rho_c - \rho_0)(\hat{a}_2)_l^m \delta(r' - a_2)$, respectively, where ρ_w is the density of seawater and ρ_c is the core density. The results are

$$(\hat{\chi}_1)_l^m(r) = -\frac{4\pi G a_1}{2l+1} (\rho_0 - \rho_w) \left(\frac{r}{a_1}\right)^l (\hat{a}_1)_l^m \qquad [130a]$$

$$(\hat{\chi}_2)_l^m(r) = -\frac{4\pi G a_2}{2l+1} (\rho_c - \rho_0) \left(\frac{a_2}{r}\right)^{l+1} (\hat{a}_2)_l^m \qquad [130b]$$

The boundary deformations $\hat{a}_n$ are determined from the principle that the normal stress must be continuous across the deformed surfaces $r = a_n + \hat{a}_n \equiv r_n$. This is equivalent to the requirement that the nonhydrostatic normal stress $\hat{\sigma}_{rr}$ acting on the reference surfaces $r = a_n$ be equal to the weight of the topography there. Expanded in spherical harmonics, this condition is

$$-\hat{p}_l^m(a_n) + 2\eta \frac{d\hat{w}_l^m}{dr}(a_n) = -g_0 \Delta\rho_n (\hat{a}_n)_l^m \qquad [131]$$

where $\hat{p}$ is the nonhydrostatic pressure, $\hat{w}$ is the radial velocity, $\Delta\rho_1 = \rho_0 - \rho_w$, and $\Delta\rho_2 = \rho_0 - \rho_c$. An expression for $\hat{p}$ in terms of the poloidal scalar is obtained by integrating the $\mathbf{e}_\theta$-component of the momentum eqn [115a] and expanding the result in spherical harmonics:

$$\hat{p}_l^m = -\eta \frac{d}{dr}(r D_l \mathcal{P}_l^m) - \rho_0 \hat{\chi}_l^m \qquad [132]$$

Substituting eqn [132] and $\hat{w}_l^m = -l(l+1)\mathcal{P}_l^m / r$ into eqn [131] and applying the boundary conditions [121], we obtain

$$(\hat{a}_n)_l^m = \Delta\rho_n^{-1} \left[X_l^m(a_n) - \frac{\rho_0}{g_0} \hat{\chi}_l^m(a_n)\right] \qquad [133a]$$

$$X_l^m(r) = \frac{\eta}{g_0} \left[-r \frac{d^3}{dr^3} + \frac{3l(l+1)}{r} \frac{d}{dr}\right] \mathcal{P}_l^m(r) \qquad [133b]$$

Now by substituting eqn [130] into eqn [126] and using eqn [133a], we obtain the following expression for the total gravitational potential:

$$\hat{\chi}_l^m(r) = (\hat{\chi}_0)_l^m(r) - \frac{4\pi a_1 G}{2l+1} \left(\frac{r}{a_1}\right)^l \left[X_l^m(a_1) - \frac{\rho_0}{g_0} \hat{\chi}_l^m(a_1)\right] + \frac{4\pi a_2 G}{2l+1} \left(\frac{a_2}{r}\right)^{l+1} \left[X_l^m(a_2) - \frac{\rho_0}{g_0} \hat{\chi}_l^m(a_2)\right] \qquad [134]$$

The boundary potentials $\hat{\chi}_l^m(a_1)$ and $\hat{\chi}_l^m(a_2)$ are determined by solving the coupled equations obtained by evaluating eqn [134] at $r = a_1$ and $r = a_2$ and are then eliminated from eqn [134]. Next, the resulting equation for $\hat{\chi}_l^m(r)$ is rewritten as a convolution integral using eqns [129], [133b], [119], [120], and [125]. Finally, the result is evaluated at $r = a_1$ to obtain the surface potential

$$\hat{\chi}_l^m(a_1) = -\frac{4\pi a_1 G}{2l+1} \int_{a_2}^{a_1} G_l(r') \hat{\rho}_l^m(r') dr' \qquad [135]$$

where the Green's function or kernel G_l is

$$G_l(r) = \left(1 - K_1 + K_2 K_r \beta^{2l+1}\right)^{-1} \left(\frac{r}{a_1}\right)^{l+2} \left\{1 - K_r \beta \left(\frac{a_2}{r}\right)^{2l+1} - E_1 \left(1 - K_r \beta^{2l+2}\right) \left(\frac{a_1}{r}\right)^{l+2} - E_2 (1 - K_r \beta) \left(\frac{a_2}{r}\right)^{l+2}\right\} \qquad [136a]$$

$$E_n = \frac{l(l+2)}{2l+1}\left(\frac{a_n}{r}\right)^{l-2}\frac{(r/a_1)^{2l-1}-(a_2/a_n)^{2l-1}}{1-\beta^{2l-1}} + \frac{l(l-1)}{2l+1}\left(\frac{a_n}{r}\right)^{l}\frac{(r/a_1)^{2l+3}-(a_2/a_n)^{2l+3}}{1-\beta^{2l+3}} \quad [136b]$$

$$K_n = \frac{4\pi a_n \rho_0 G}{(2l+1)g_0}, \quad K_r = \frac{K_1}{1+K_2}, \quad \beta = \frac{a_2}{a_1} \quad [136c]$$

Figure 10 shows $G_l(r)$ for $l=2$ and $l=10$, both with (solid lines) and without ($K_1=K_2=0$; dashed lines) self-gravitation. The kernels are negative at all radii because the negative gravitational potential of the deformed upper surface exceeds the positive contribution of the internal mass anomaly itself. The maximum potential anomaly is produced by loads in the mid-mantle when $l=2$ and in the upper mantle when $l=10$. The effect of self-gravitation is nearly a factor of two for $l=2$, but only about 10% for $l=10$.

Forte and Peltier (1987) also gave an expression analogous to eqn [136] for a mantle comprising an upper layer $r>a_1-z_{LU}$ with viscosity η_U and a lower layer $r<a_1-z_{LU}$ with viscosity η_L. The general effect of a viscosity contrast $\eta_L/\eta_U>1$ is to enhance the dynamic deformation of the CMB and reduce that of the upper surface. The reduced (negative) gravitational potential anomaly of the upper surface then counteracts less effectively the positive anomaly due the (sinking) load, with the result that $G_l(r)$ increases at all depths relative to the kernels for $\eta_L/\eta_U=1$. For $\eta_L/\eta_U=30$ and $z_{LU}=670$ km, for example, $G_{10}(r)>0$ for all r, while $G_2(r)$ is positive in the upper mantle and negative in the lower mantle (Forte and Peltier, 1987, Fig. 17).

The simplest application of the kernel approach (e.g., Forte and Peltier, 1987; Hager et al., 1985) is to start with a load function $\hat{\rho}(r,\theta,\phi)$ estimated from seismic tomography or the global distribution of subducted lithosphere and then to determine by repeated forward modeling the values of η_L/η_U and z_{LU} for which geoid and other anomalies (surface divergence, surface topography, etc.) predicted by formulas like eqn [135] best match their observed counterparts on the real Earth. A robust result of the early studies that have been confirmed by later work (e.g., Mitrovica and Forte, 2004) is that the lower mantle must be more viscous than the upper mantle by a factor $\eta_L/\eta_U \sim 10-100$. In many of these studies, the kernels were obtained using an alternative analytical technique, the propagator matrix method, discussed next.

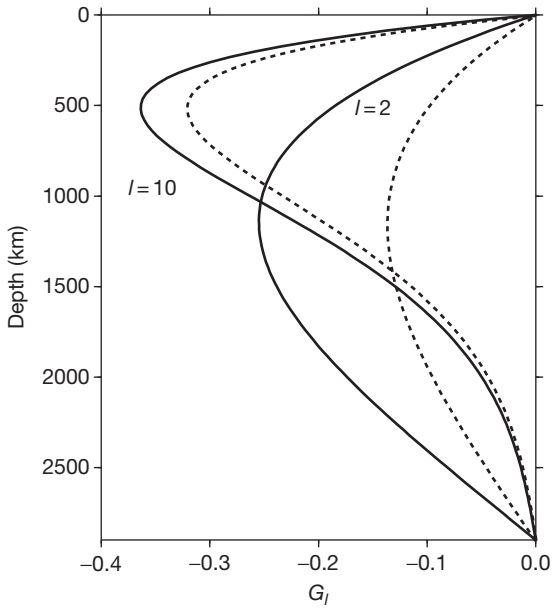


Figure 10 Gravitational potential kernels $G_l(r)$ defined by eqn [136] as functions of depth for spherical harmonic degrees $l=2$ and $l=10$, with (solid lines) and without (dashed lines) self-gravitation.

7.04.5.8.2 The propagator matrix method

The analytic Green's function approach outlined in Section 7.04.5.8.1 for a constant-viscosity mantle can in principle be extended to models comprising any number N of discrete layers with different viscosities. In practice, however, the rapidly increasing complexity of the analytical expressions limits the method to $N=2$ (Forte and Peltier, 1987). A more efficient approach for models with multiple layers is the propagator matrix method, whereby a flow solution is propagated from one layer interface to the next by simple matrix multiplication. The method is in fact applicable to any system of linear ODEs with constant coefficients of the form

$$\frac{d\mathbf{y}}{dz} = \mathbf{A}\mathbf{y} + \mathbf{b}(z) \quad [137]$$

where $\mathbf{y}(z)$ is a vector of dependent variables, $\mathbf{A}$ is a constant square matrix, and $\mathbf{b}(z)$ is an inhomogeneous vector. The general solution of eqn [137] is (Gantmacher, 1960, vol. I, p. 120, eqn [53])

$$\mathbf{y}(z) = \mathbf{P}(z,z_0)\mathbf{y}(z_0) + \int_{z_0}^{z}\mathbf{P}(z,\zeta)\mathbf{b}(\zeta)d\zeta \quad [138a]$$

$$\mathbf{P}(z,z_0) = \exp[\mathbf{A}(z-z_0)] \quad [138b]$$

where $\mathbf{y}(z_0)$ is the solution vector at the reference point $z=z_0$ and $\mathbf{P}(z,z_0)$ is the propagator matrix. The form of the solution [138] is identical to that for a scalar variable $y(z)$, except that the argument of the exponential in eqn [138b] is now a matrix rather than a scalar quantity. Analytical expressions for functions of matrices such as eqn [138b] are given by Gantmacher (1960, vol. I, pp. 95–110).

As an illustration, we determine the propagator matrix for a poloidal flow driven by internal density anomalies in a self-gravitating spherical shell with radially variable viscosity $\eta(r)$ and laterally averaged density $\rho(r)$ (Hager and O'Connell, 1981; Richards and Hager, 1984). As in Section 7.04.5.8.1, we suppose that the driving density anomaly is $\hat{\rho}_l^m(r)Y_l^m(\theta,\phi)$, where Y_l^m is a surface spherical harmonic.

The first step is to transform the equations governing the flow in the mantle to the canonical form [137], where $\mathbf{A}$ – to repeat – is a constant matrix. The standard Stokes equations [43a]–[43e] for our model mantle do not have this form for three reasons: (i) the fluid properties $\eta(r)$ and $\rho(r)$ vary with radius, (ii) the expression for the differential operator ∇ in spherical coordinates involves the scale factor r^{-1}, and (iii) a system of first-order equations cannot be written in terms of the primitive variables u_i and p. Difficulty (i) is circumvented by dividing the mantle into N discrete layers $n=1,2,\ldots,N$, in each of which the viscosity η_n and the density ρ_n are constant. Difficulty (ii) is overcome by using a transformed radial variable $z=\ln(r/a_1)$, where a_1 is the outer radius of the mantle. Finally, one circumvents (iii) by using the independent variables

$$\mathbf{y} = \left[\hat{u}, \hat{v}, \frac{r\hat{\sigma}_{rr}}{\eta_0}, \frac{r\hat{\sigma}_{r\theta}}{\eta_0}, \frac{\rho_0 r\hat{\chi}}{\eta_0}, \frac{\rho_0 r^2 \partial_r \hat{\chi}}{\eta_0}\right]^{\mathcal{T}} \quad [139]$$

where $\hat{u}$ and $\hat{v}$ are the $\mathbf{e}_r$- and $\mathbf{e}_\theta$-components, respectively, of the velocity, η_0 and ρ_0 are reference values of the viscosity and density, respectively, and the spherical harmonic dependence of each variable ($\hat{u}$, $\hat{\sigma}_{rr}$, $\hat{\chi}$, $\partial_r\hat{\chi} \propto Y_l^m$; $\hat{v}$, $\hat{\sigma}_{r\theta} \propto \partial_\theta Y_l^m$) has been suppressed for clarity. The Stokes equations within each layer n then take the form [137] with

$$\mathbf{A} = \begin{bmatrix} -2 & L & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1/\eta^\star & 0 & 0 \\ 12\eta^\star & -6L\eta^\star & 1 & L & 0 & \rho^\star \\ -6\eta^\star & 2(2L-1)\eta^\star & -1 & -2 & \rho^\star & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & L & 0 \end{bmatrix} \quad [140a]$$

$$\mathbf{b} = \left[0, 0, r^2 g_0(r)\hat{\rho}(r)/\eta_0, 0, 0, -4\pi r^3 G\rho_0\hat{\rho}(r)/\eta_0\right]^{\mathrm{T}} \quad [140b]$$

where $\hat{\rho} \equiv \hat{\rho}_l^m$, $L = l(l+1)$, $\eta^\star = \eta_n/\eta_0$, and $\rho^\star = \rho_n/\rho_0$. Further simplification is achieved by recasting the density anomaly $\hat{\rho}(r)$ as a sum of equivalent surface density contrasts ξ_n (units of kg m^{-2}) localized at the midpoints r_n of the layers, according to

$$\hat{\rho}(r) = \sum_n \delta(r - r_n)\xi_n \quad [141]$$

In practice, ξ_n is different in each layer, while η_n and ρ_n may be constant over several adjacent layers. For a group of M such adjacent layers bounded by the depths z and z_0, the solution [138a] now takes the approximate form

$$\mathbf{y}(z) = \mathbf{P}(z, z_0)\mathbf{y}(z_0) + \sum_{m=1}^{M} \mathbf{P}(z, z_m)\mathbf{b}_m \quad [142a]$$

$$\mathbf{b}_m = \left[0, 0, r_m g_0(r_m)\xi_m/\eta_0, 0, 0, -4\pi r_m^2 G\rho_0\xi_m/\eta_0\right]^{\mathrm{T}} \quad [142b]$$

where $z_m = \ln(r_m/a_1)$. To use [142], one must apply boundary conditions at the CMB and at the Earth's surface, taking into account the dynamic topography of these boundaries (Richards and Hager, 1984, Appendix 1).

An analytical expression for $\mathbf{P}(z, z_0)$ can be written in terms of the minimal polynomial $\psi(\lambda)$ of $\mathbf{A}$. For the matrix [140a], $\psi(\lambda)$ is identical to the characteristic polynomial, and is

$$\psi(\lambda) = \prod_{i=1}^{4} (\lambda - \lambda_i)^{m_i} \quad [143a]$$

$$\lambda_1 = l+1, \quad \lambda_2 = -l, \quad \lambda_3 = l-1, \quad \lambda_4 = -l-2 \quad [143b]$$

$$m_1 = m_2 = 2, \quad m_3 = m_4 = 1 \quad [143c]$$

where $\lambda_1 - \lambda_4$ are the four distinct eigenvalues of $\mathbf{A}$. Following Gantmacher (1960, I, pp. 95-102),

$$\mathbf{P}(z, z_0) = \sum_{k=1}^{4} \sum_{j=1}^{m_k} \alpha_{kj} (\mathbf{A} - \lambda_k \mathbf{I})^{j-1} \Psi_k \quad [144]$$

where

$$\alpha_{kj} = \frac{1}{(j-1)!} \frac{\mathrm{d}^{j-1}}{\mathrm{d}\lambda^{j-1}} \left[\frac{\exp\lambda(z - z_0)}{\psi_k(\lambda)}\right]_{\lambda=\lambda_k} \quad [145]$$

$$\psi_k(\lambda) = \frac{\psi(\lambda)}{(\lambda - \lambda_k)^{m_k}}, \quad \Psi_k = \prod_{\substack{i=1 \\ i \neq k}}^{4} (\mathbf{A} - \lambda_i \mathbf{I})^{m_i} \quad [146]$$

and $\mathbf{I}$ is the identity matrix.

7.04.6 Elasticity and Viscoelasticity

Fluid convection with a period $\tau \sim 10^{15}$s is the extreme limit of a spectrum of deformational processes in the mantle spanning an enormous range of timescales, including postglacial rebound ($\tau \sim 10^{11}$s), the Chandler wobble ($\tau \sim 4 \times 10^7$s), and elastic free oscillations ($\tau \sim 10 - 10^3$s). All these processes can be understood by regarding the mantle as a viscoelastic body that deforms as a fluid when $\tau \gg \tau_M$ and as an elastic solid when $\tau \ll \tau_M$, where τ_M is the Maxwell time of the material. τ_M is just the ratio of the viscosity of the material to its elastic shear modulus, and is of the order of a few hundred years ($\sim 10^{10}$s) in the mantle.

Because the theory of viscous flow is valid only at very long-periods ($\tau \gg \tau_M$), other rheological models are required to understand phenomena with shorter periods. The two most commonly used models are the linear elastic solid (for short periods, $\tau \ll \tau_M$) and the linear Maxwell solid (for intermediate periods, $\tau \sim \tau_M$). The reader is referred to Chapter 7.02 for a more extensive discussion of these models. Here, we focus on the mathematical analogies (correspondence principles) among the viscous, elastic, and viscoelastic models and the powerful analytical techniques these analogies make possible.

7.04.6.1 Correspondence Principles

Stokes (1845) and Rayleigh (1922) demonstrated the existence of a mathematical correspondence (the Stokes–Rayleigh analogy) between small incompressible deformations of an elastic solid and slow flows of a viscous fluid. As discussed in Chapter 7.02, the constitutive law for a linear (Hookean) elastic solid is

$$\sigma_{ij} = K\varepsilon_{kk}\delta_{ij} + 2\mu\left(\varepsilon_{ij} - \frac{1}{3}\varepsilon_{kk}\delta_{ij}\right) \quad [147]$$

where σ_{ij} is the stress tensor, $\varepsilon_{ij} \equiv (\partial_i\zeta_j + \partial_j\zeta_i)/2$ is the linearized strain tensor, ζ_i is the displacement vector, K is the bulk modulus, and μ is the shear modulus. Alternatively, eqn [147] can be written in terms of Young's modulus E and Poisson's ratio ν, which are related to K and μ by

$$E = \frac{9K\mu}{3K + \mu}, \quad \nu = \frac{3K - 2\mu}{2(3K + \mu)} \quad [148]$$

An incompressible elastic solid corresponds to the limits $K/E \to \infty$ and $\nu \to 1/2$ and an incompressible deformation to the limit $\Delta \to 0$. As these limits are approached, however, the product $-K\Delta$ tends to a finite value, the pressure p. Equation [147] then becomes

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\varepsilon_{ij} \quad [149]$$

which is identical to the constitutive relation [43f] for a viscous fluid if the shear modulus μ and the displacement ζ_i are

replaced by the viscosity η and the velocity u_i, respectively. Consequently, the equations of (incompressible) Stokes flow can be obtained from those of (compressible) linear elasticity via the transformation

$$(\zeta_i, E, \nu) \rightarrow (u_i, 3\eta, 1/2) \qquad [150]$$

The transformation eqn [150] will be used in **Section 7.04.8.4** to obtain equations governing the slow deformation of thin viscous shells.

In principle, the Stokes–Rayleigh analogy between linear elasticity and slow viscous flow applies not only to the governing equations themselves but also to the solutions of BVPs for those equations. In that case, however, the analogy requires that the boundary conditions of the elastic and viscous problems be properly related. For example, imposed boundary velocities in the viscous problem must correspond to imposed boundary displacements in the elastic one. However, while viscous flows driven by boundary velocities are relatively common in the geophysical literature, elastic deformations driven by prescribed boundary displacements are not. This limits the utility of the Stokes–Rayleigh analogy as a tool for solving viscous flow problems.

A second useful correspondence principle relates problems in linear viscoelasticity and linear elasticity. Although this principle is valid for any linear viscoelastic body (Biot, 1954), its geophysical application is usually limited to the special case of a linear Maxwell solid, for which the constitutive relation is (e.g., Peltier, 1974)

$$\dot{\sigma}_{ij} + \frac{\mu}{\eta}\left(\sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij}\right) = 2\mu\left(\dot{\varepsilon}_{ij} - \frac{1}{3}\dot{\varepsilon}_{kk}\delta_{ij}\right) + K\dot{\varepsilon}_{kk}\delta_{ij} \qquad [151]$$

where dots denote time derivatives. Transforming eqn [151] into the frequency domain using a Laplace transform as described in **Chapter 7.02**, we obtain

$$\overline{\sigma}_{ij} = K\overline{\varepsilon}_{kk}\delta_{ij} + 2\overline{\mu}(s)\left(\overline{\varepsilon}_{ij} - \frac{1}{3}\overline{\varepsilon}_{kk}\delta_{ij}\right) \qquad [152a]$$

$$\overline{\mu}(s) = \frac{\mu s}{s + \mu/\eta} \qquad [152b]$$

where s is the complex frequency and overbars on $\sigma_{ij}, \varepsilon_{ij}$ and ε_{kk} denote Laplace transforms. Equation [152] is identical in form to Hooke's law [147] for an elastic solid, but with a frequency-dependent shear modulus $\overline{\mu}$. Accordingly, any viscoelastic problem can be reduced to an equivalent elastic problem, which can be solved and then inverse Laplace transformed back into the time domain to yield the solution of the original viscoelastic problem. We now turn to one of the principle geophysical applications of this procedure.

7.04.6.2 Surface Loading of a Stratified Elastic Sphere

The fundamental problem of postglacial rebound is to determine the response of a radially stratified Maxwell Earth to a time-dependent surface load $P(\theta, \phi, t)$ that represents the changing global distribution of glacial ice and meltwater. Formally, this problem can be solved by convolving the load distribution with a Green's function that describes the Earth's response to an impulsive point load applied at time t at a point $\mathbf{r}$ on the Earth's surface and then immediately removed. Now, an impulse function in the time domain corresponds to a constant in the frequency domain. According to the correspondence principle, therefore, our problem reduces to that of determining the deformation of a stratified elastic sphere by a static point load (Farrell, 1972; Longman, 1962; Peltier, 1974). Our derivation follows Peltier (1995).

The general equations governing the deformation of a self-gravitating elastic sphere are (Backus, 1967)

$$0 = \nabla\cdot\hat{\boldsymbol{\sigma}} + \mathbf{g}_0\hat{\rho} - \rho_0\nabla\hat{\chi} - \nabla(\rho_0 g_0\hat{u}) \qquad [153a]$$

$$\nabla^2\hat{\chi} = 4\pi G\hat{\rho} \qquad [153b]$$

where $\hat{u}$ is the radial component of the displacement vector $\hat{\boldsymbol{\zeta}} \equiv \hat{u}\mathbf{e}_r + \hat{v}\mathbf{e}_\theta + \hat{w}\mathbf{e}_\phi$ and the other symbols are the same as in the analogous eqns [115a] and [115b] for slow viscous flow. There are three significant differences between the two sets of equations. First, the stress tensor $\hat{\boldsymbol{\sigma}}$ in eqn [153a] is related to the displacement vector $\hat{\boldsymbol{\zeta}}$ by the elastic constitutive law eqn [152] with a frequency-dependent shear modulus. Second, because deformation in the elastic case is driven by surface loading rather than by internal density anomalies, the perturbation density $\hat{\rho} \equiv -\nabla\cdot\left(\rho_0\hat{\boldsymbol{\zeta}}\right)$ in eqn [153] is determined entirely by the requirement of mass conservation in the deformed solid. Third, eqn [153a] contains an additional term $-\nabla(\rho_0 g_0\hat{u})$ that corrects for the fact that the strain tensor that appears in the elastic constitutive law is at a fixed material particle and not at a fixed point in space (Backus, 1967, p. 96).

The reduction of the governing eqn [153] to solvable form proceeds as for the viscous flow equations in **Section 7.04.5.8.2**, but with a few significant differences. As in the viscous case, the goal is to reduce eqn [153] to a sixth-order system of ODEs of the form

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}r} = \mathbf{A}\mathbf{y} \qquad [154]$$

where $\mathbf{y}$ is the unknown vector and $\mathbf{A}$ is a 6×6 matrix. Equation [154] contains no inhomogeneous vector $\mathbf{b}$ like the one in the viscous equations [137], because the loads in the elastic problem are applied at the boundaries rather than internally. Another important point is that the reference profiles of density $\rho_0(r)$ and the elastic moduli $\mu(r)$ and $K(r)$ that appear in the equations are essentially continuous functions, given a priori by a seismological reference model (e.g., PREM; Dziewonski and Anderson, 1981). Consequently, a discrete-layer solution method like the propagator matrix technique is not practical, and it is therefore superfluous to reduce eqn [154] to constant-coefficient form via the transformation $z = \ln(r/a_1)$.

We turn now to the definition of the variables $\mathbf{y}$. By symmetry, the azimuthal deformation $\hat{w}$ is zero, and the components $\hat{u}$ and $\hat{v}$ can depend only on the radius r and the angular distance θ from the point load. The field variables can therefore be expanded as

$$\hat{\boldsymbol{\zeta}} = \sum_{l=0}^{\infty}\left[\hat{u}_l(r,s)P_l(\cos\theta)\mathbf{e}_r + \hat{v}_l(r,s)\partial_\theta P_l(\cos\theta)\mathbf{e}_\theta\right] \qquad [155a]$$

$$\hat{\chi} = \sum_{l=0}^{\infty}\hat{\chi}_l(r,s)P_l(\cos\theta) \qquad [155b]$$

where $P_l(\cos\theta)$ are the Legendre polynomials. Now, for each angular degree l, let

$$\mathbf{y} = [\hat{u}, \hat{v}, \hat{\sigma}_{rr}, \hat{\sigma}_{r\theta}, \hat{\chi}, \partial_r \hat{\chi} + (l+1)\hat{\chi}/r + 4\pi G\rho_0 \hat{u}]^{\mathrm{T}} \quad [156]$$

where the subscript l on each variable has been suppressed for simplicity. Apart from the factors of r, the definition of y_6, and the replacement of velocities by displacements, eqn [156] is identical to its viscous analogue [139]. With the choice [156], the matrix $\mathbf{A}$ in eqn [154] is

$$\mathbf{A} = \begin{bmatrix} -\frac{2\lambda}{\beta r} & \frac{L\lambda}{\beta r} & \frac{1}{\beta} & 0 & 0 & 0 \\ -r^{-1} & r^{-1} & 0 & \overline{\mu}^{-1} & 0 & 0 \\ \frac{4}{r}\left(\frac{\gamma}{r} - \rho_0 g_0\right) & -\frac{L}{r}\left(\frac{2\gamma}{r} - \rho_0 g_0\right) & -\frac{4\overline{\mu}}{\beta r} & \frac{L}{r} & -\frac{\rho_0 M}{r} & \rho_0 \\ \frac{1}{r}\left(\rho_0 g_0 - \frac{2\gamma}{r}\right) & \frac{L(\gamma+\overline{\mu}) - 2\overline{\mu}}{r^2} & -\frac{\lambda}{\beta r} & -\frac{3}{r} & \frac{\rho_0}{r} & 0 \\ -4\pi G\rho_0 r & 0 & 0 & 0 & -\frac{M}{r} & 1 \\ -\frac{4\pi M G\rho_0}{r} & \frac{4\pi L G\rho_0}{r} & 0 & 0 & 0 & \frac{l-1}{r} \end{bmatrix} \quad [157]$$

where $L = l(l+1)$, $M = l+1$, $\beta = K + 4\overline{\mu}/3$, $\gamma = 3K\overline{\mu}/\beta$, $\lambda = K - 2\overline{\mu}/3$, and $\overline{\mu}$ is defined by eqn [152b]. The system [154] can be solved by a standard numerical method for two-point boundary-value problems (e.g., shooting) subject to appropriate boundary conditions. Inversion of the resulting solution back into the time domain and its subsequent convolution with a given time-dependent surface load function are then performed numerically. The use of this procedure to infer the mantle viscosity profile $\eta(r)$ is reviewed by Peltier (1995).

7.04.7 Boundary-Layer Theory

7.04.7.1 The Boundary-Layer Approximation

Many geophysical flows occur in layers or conduits whose length greatly exceeds their thickness: examples include TBLs, subducted oceanic lithosphere, mantle plume stems, and gravity currents of buoyant plume material spreading beneath the lithosphere. In all these cases, the gradients of the fluid velocity and/or temperature across the layer greatly exceed the gradients along it, a fact that can be exploited to simplify the governing equations substantially. The classic example of this approach is boundary-layer theory, which describes the flow in layers whose thickness is controlled by a balance of diffusion and advection.

A boundary layer is defined as a thin region in a flow field, usually adjoining an interface or boundary, where the gradients of some quantity transported by the fluid (e.g., vorticity, temperature, or chemical concentration) are large relative to those elsewhere in the flow. Physically, boundary layers arise when the boundary acts as a source of the transported quantity, which is then prevented from diffusing far from the boundary by strong advection. Boundary layers thus occur when $UL/D \gg 1$, where U and L are characteristic velocity and length scales for the flow and D is the diffusivity of the quantity in question. In classical boundary-layer theory, the transported quantity is vorticity, the relevant diffusivity is the kinematic viscosity ν, and boundary layers form when the Reynolds number $\mathrm{Re} \equiv UL/\nu \gg 1$. Such boundary layers do not occur in the mantle, where $\mathrm{Re} \sim 10^{-20}$. However, because the thermal diffusivity $\kappa \sim 10^{-23}\nu$, the Péclet number $UL/\kappa \gg 1$ for typical mantle flows, implying that TBLs will be present.

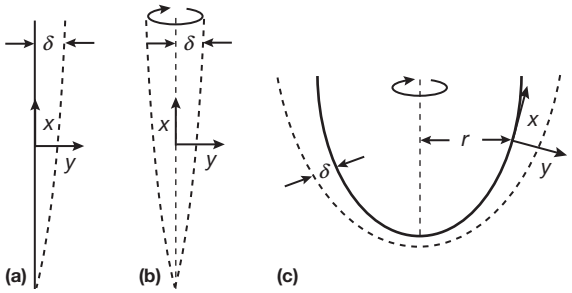


Figure 11 Typical geometries for boundary-layer flows: (a) 2-D flow, (b) an axisymmetric plume, and (c) axisymmetric flow along a surface of revolution. Impermeable boundaries are shown by heavy lines, axes of symmetry by light dashed lines, and the edges of BLs by heavy dashed lines. The coordinates x and y are parallel to and normal to the boundary (or symmetry axis), respectively; $\delta(x)$ is the thickness of the BL, and $r(x)$ is the radius of the surface of revolution.

Although BLs are 3-D structures in general, nearly all BL models used in geodynamics involve one of the three simple geometries shown in Figure 11: 2-D flow (a), an axisymmetric plume (b), and axisymmetric flow along a surface of revolution (c). Let x and y be the coordinates parallel to and normal to the boundary (or symmetry axis), respectively; u and v be the corresponding velocity components; and $\delta(x)$ be the thickness of the BL.

The fundamental hypothesis of BL theory is that gradients of the transported quantity (heat in this case) along the BL are much smaller than those across it. Diffusion of heat along the layer is therefore negligible. As an illustration, consider the simplest case of a thermal BL in steady flow with negligible viscous dissipation of energy. The BL forms of the continuity and energy equations are then

$$u_x + v_y = 0 \quad [158a]$$

$$uT_x + vT_y = \kappa T_{yy} \quad [158b]$$

for 2-D flow,

$$u_x + \gamma^{-1}(\gamma v)_y = 0 \quad [159a]$$

$$uT_x + vT_y = \kappa\gamma^{-1}(\gamma T_y)_y \quad [159b]$$

for an axisymmetric plume, and

$$(ru)_x + rv_y = 0 \quad [160a]$$

$$uT_x + vT_y = \kappa T_{yy} \quad [160b]$$

for a surface of revolution. As noted in Section 7.04.3.3, the terms uT_x and vT_y are of the same order.

7.04.7.2 Solution of the Boundary-Layer Equations Using Variable Transformations

A powerful technique for solving problems involving BLs is the use of variable transformations. One of the most important of these is that of von Mises (1927), which transforms the 2-D BL equations into the classical heat conduction equation. A second useful transformation, due to Mangler (1948), relates

the structure of an axisymmetric BL on a surface of revolution to that of a 2-D BL on a flat surface. After introducing both transformations, we will show how they can be used together to obtain a solution for the heat transfer from a hot sphere moving in a viscous fluid (Figure 1(a)).

7.04.7.2.1 The von Mises transformation

The essential trick involved in this transformation is to use the stream function ψ instead of y as the transverse coordinate in the BL. Denoting the streamwise coordinate by a new symbol $s \equiv x$ for clarity, we transform the derivatives in eqn [158b] using the chain rule as

$$\frac{\partial T}{\partial x} = \frac{\partial T}{\partial s}\frac{\partial s}{\partial x} + \frac{\partial T}{\partial \psi}\frac{\partial \psi}{\partial x} \equiv \frac{\partial T}{\partial s} - v\frac{\partial T}{\partial \psi} \quad [161]$$

$$\frac{\partial T}{\partial y} = \frac{\partial T}{\partial s}\frac{\partial s}{\partial y} + \frac{\partial T}{\partial \psi}\frac{\partial \psi}{\partial y} \equiv u\frac{\partial T}{\partial \psi} \quad [162]$$

where the stream function is defined according to the convention $(u, v) = (\psi_{y'} - \psi_x)$. Equation [158b] then becomes

$$\frac{\partial T}{\partial s} = \kappa\frac{\partial}{\partial \psi}\left(u\frac{\partial T}{\partial \psi}\right) \quad [163]$$

Equation [163] takes still simpler forms if the surface $y = \psi = 0$ is either free-slip or rigid. Near a free-slip surface, $u \approx U(s)$ is constant across the BL. Upon introducing a new downstream coordinate τ such that $\mathrm{d}\tau = U(s)\mathrm{d}s$, eqn [163] becomes

$$\frac{\partial T}{\partial \tau} = \kappa\frac{\partial^2 T}{\partial \psi^2} \quad [164]$$

which is just the classical equation for diffusion of heat in a medium with constant thermal diffusivity κ. The units of the time-like variable τ, however, are now those of diffusivity ($\mathrm{m}^2\ \mathrm{s}^{-1}$) rather than of time. Near a rigid surface, $u = yf(s)$ and $\psi = y^2 f(s)/2$, implying $u \approx (2\psi f)^{1/2}$, where $f(s)$ is arbitrary. Substituting this result into eqn [163] and introducing a new downstream variable τ such that $\mathrm{d}\tau = (2f)^{1/2}\mathrm{d}s$, we obtain

$$\frac{\partial T}{\partial \tau} = \kappa\frac{\partial}{\partial \psi}\left(\psi^{1/2}\frac{\partial T}{\partial \psi}\right) \quad [165]$$

which describes the diffusion of heat in a medium with a position-dependent thermal diffusivity $\kappa\psi^{1/2}$ as a function of a time-like variable τ having units of $\mathrm{m\ s}^{-2}$.

Of special interest are the self-similar solutions of eqns [164] and [165] that exist when the wall temperature T_0 is constant (say, $=\Delta T$) and the upstream temperature profile is $T(0, \psi) = 0$. The solution of the free-slip surface eqn [164] has the form $T = \Delta T F(\eta)$, where $\eta = \psi/\delta(\tau)$. Separating variables, we obtain $-F''/\eta F' = \delta\dot{\delta}/\kappa = \lambda^2$ (constant), which when solved subject to the conditions $F(0) - 1 = F(\infty) = \delta(0) = 0$ yields

$$T = \Delta T\,\mathrm{erfc}\left(\frac{\psi}{2\sqrt{\kappa\tau}}\right) \quad [166]$$

The solution of the rigid-surface eqn [165] has the form $T = \Delta T F(\eta)$, where $\eta = \sqrt{\psi}/\delta(\tau)$. Separating variables, we find $-F''/4\eta^2 F' = \delta^2\dot{\delta}/\kappa = \lambda^2$ (constant), which when solved subject to $F(0) - 1 = F(\infty) = \delta(0) = 0$ with $\lambda^2 = 1/3$ for convenience gives

$$\frac{T}{\Delta T} = 1 - \frac{3}{\Gamma(1/3)}\left(\frac{4}{9}\right)^{1/3}\int_0^{\eta}\exp\left(-\frac{4}{9}x^3\right)\mathrm{d}x, \quad \eta = \frac{\psi^{1/2}}{(\kappa\tau)^{1/3}} \quad [167]$$

where Γ is the gamma function. Later in the text, we will see how the solution [166] can be applied to heat transfer from a moving sphere (Section 7.04.7.2.2) and to steady cellular convection (Section 7.04.9.3.3).

7.04.7.2.2 Mangler's transformation

Mangler (1948) showed that the equations governing the axisymmetric BL on a surface of revolution with radius $r(x)$ (Figure 11) are related to the 2-D BL equations by the variable transformation

$$\bar{x} = \int_0^x (r/L)^2\mathrm{d}x, \ \bar{y} = \frac{r}{L}y, \ \bar{u} = u, \ \bar{v} = \frac{L}{r}\left(v + \frac{r'}{r}yu\right), \ \bar{\psi} = L^{-1}\psi \quad [168]$$

where the variables with and without overbars are those of the 2-D and the axisymmetric flows, respectively, and L is an arbitrary constant length scale. The transformation [168], which applies equally to vorticity and thermal BLs, allows solutions of the 2-D BL equations to be transformed directly into solutions of the axisymmetric BL equations on a surface of revolution. To illustrate, we determine the heat flow from a sphere of radius a with a free-slip surface and excess temperature ΔT in a stream of viscous fluid with velocity $-U\mathbf{e}_z$ far from the sphere (Figure 1). The starting point is the Cartesian BL solution [166], which in terms of the barred variables is

$$T = \Delta T\,\mathrm{erfc}\left(\frac{\bar{\psi}}{2\sqrt{\kappa\bar{\tau}}}\right), \quad \bar{\tau}(\bar{x}) = \int_0^{\bar{x}}\bar{u}\mathrm{d}\bar{x} \quad [169]$$

For a sphere, $r(x) = a\sin(x/a) \equiv a\sin\theta$, and $L = a$ is the natural choice. The Stokes–Hadamard solution [64] with $\gamma = 0$ gives $\psi \approx (1/2)Ua y\sin^2\theta$, and $u \approx (1/2)U\sin\theta$, and eqn [168] implies

$$\bar{\psi} = \frac{1}{2}Uy\sin^2\theta, \quad \bar{x} = \frac{1}{2}x - \frac{a}{4}\sin\frac{2x}{a} \quad [170]$$

The stretched downstream variable τ is therefore

$$\tau = \int_0^{\bar{x}}\bar{u}\mathrm{d}\bar{x} = \int_0^x u\frac{\mathrm{d}\bar{x}}{\mathrm{d}x}\mathrm{d}x = \frac{2}{3}aU(2 + \cos\theta)\sin^4\frac{\theta}{2} \quad [171]$$

Substitution of eqns [170] and [171] into eqn [169] yields the temperature $T(y, \theta)$ everywhere in the BL, and the corresponding local Nusselt number is

$$\mathcal{N}(\theta) \equiv -\frac{a}{\Delta T}\frac{\partial T}{\partial y}(y = 0) = \left(\frac{3}{2\pi}\right)^{1/2}\frac{1 + \cos\theta}{(2 + \cos\theta)^{1/2}}\mathrm{Pe}^{1/2} \quad [172]$$

where $\mathrm{Pe} = Ua/\kappa$ is the Péclet number.

The result [172], together with our previous treatments of the hot sphere in Sections 7.04.3.1, 7.04.3.2, and 7.04.3.3, shows that BL theory represents a third stage in a hierarchy of techniques (dimensional analysis, scaling analysis, and BL theory) that give progressively more detailed information about the structure of the solution in the asymptotic limit of negligible inertia ($\mathrm{Re} \ll 1$) and $\mathrm{Pe} \to \infty$. Table 3 summarizes the local Nusselt number $\mathcal{N}(\theta)$ for a free-slip sphere predicted by each of the three techniques. Note that the cost of the increasing precision of the results is a decreasing range of validity.

Table 3 Heat transfer from a hot sphere for $\mathrm{Re} \ll 1$: Analytical predictions

Technique	*Local Nusselt number* $\mathcal{N}$	*Validity*
Dimensional analysis	$\mathrm{fct}(\mathrm{Pe}, \theta)$	Universal
Scaling analysis	$\mathrm{Pe}^{1/2}\mathrm{fct}(\theta)$	$\mathrm{Pe} \gg 1$
BL theory	$\left(\frac{3}{2\pi}\right)^{1/2} \frac{1+\cos\theta}{(2+\cos\theta)^{1/2}} \mathrm{Pe}^{1/2}$	$\mathrm{Pe} \gg 1, \quad \theta < \frac{\pi}{2}$

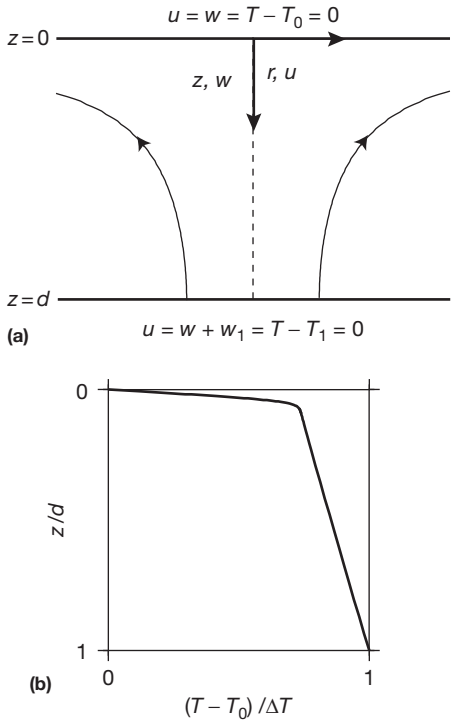


Figure 12 Model for the temperature distribution in a steady stagnation-point flow beneath a rigid lithosphere. (a) The base of the lithosphere $z=0$ is at temperature T_0, and a uniform upward vertical velocity $w_1 > 0$ is imposed at a depth $z=d$ where the temperature is T_1. The curved lines with arrows are streamlines. (b) Temperature as a function of depth for $\epsilon = 0.0001$, $\beta = 0.1$, and $\lambda = 2.0$.

7.04.7.3 The Method of Matched Asymptotic Expansions

As noted earlier (Section 7.04.5.7), the MMAE is a powerful method for solving problems where the field variables exhibit distinct regions characterized by very different length scales. The method is particularly well suited for BLs, whose characteristic thickness δ is much smaller than the scale L of the flow outside the BL. As an illustration of the method, consider a simple axisymmetric stagnation-point flow model for the steady temperature distribution in a plume upwelling beneath a rigid lithosphere (Figure 12), in which fluid with temperature T_1 and upward vertical velocity $-w_1$ at a depth $z=d$ ascends toward a rigid surface with temperature $T=T_0$. If viscous dissipation of energy is negligible, the pressure is (nearly) hydrostatic, and all physical properties of the fluid are constant, then $T(x)$ satisfies (see Chapter 7.02)

$$\mathbf{u}\cdot\nabla T - \frac{g\alpha}{c_p}\mathbf{u}\cdot\mathbf{e}_z T = \kappa\nabla^2 T \qquad [173]$$

where α is the thermal expansion coefficient and c_p is the specific heat at constant pressure. The three terms in eqn [173] represent advection of temperature gradients, adiabatic decompression, and thermal diffusion, respectively.

Let $\mathbf{u} = \mathrm{u}(r,z)\mathbf{e}_r + w(r,z)\mathbf{e}_z$. Because $w(r,d) \equiv -w_1$ is constant, $w = w(z)$, whence the continuity equation implies $u = -rw'(z)/2$. Equation [173] therefore admits a 1-D solution $T=T(z)$ that satisfies

$$wT' - \frac{g\alpha}{c_p}wT = \kappa T'' \qquad [174]$$

where primes denote d/dz. By substituting $w=w(z)$ and $u=-rw\prime(z)/2$ into the constant-viscosity Stokes equations in cylindrical coordinates and solving the resulting equation for $w(z)$ subject to the boundary conditions shown in Figure 12, we find

$$\frac{w}{w_1} = 2\left(\frac{z}{d}\right)^3 - 3\left(\frac{z}{d}\right)^2 \qquad [175]$$

Upon introducing dimensionless variables $\widetilde{z} = z/d$ and $\widetilde{T} = (T-T_0)/(T_1-T_0) \equiv (T-T_0)/\Delta T$ and then immediately dropping the tildes, eqn [173] together with eqn [175] and the boundary conditions on T become

$$\left(2z^3 - 3z^2\right)[T' - \beta(T+\lambda)] = \epsilon T'' \qquad [176a]$$

$$T(0) = T(1) - 1 = 0 \qquad [176b]$$

where

$$\epsilon = \frac{\kappa}{dw_1} \equiv \mathrm{Pe}^{-1}, \quad \beta = \frac{g\alpha d}{c_p}, \quad \lambda = \frac{T_0}{T_1 - T_0} \qquad [177]$$

We wish to solve eqn [176] in the limit $\epsilon \to 0$ ($\mathrm{Pe} \to \infty$), assuming for simplicity that $\beta = O(1)$ and $\lambda = O(1)$. However, note that ϵ appears in eqn [176a] as the coefficient of the most highly differentiated term. We therefore cannot simply set $\epsilon = 0$ in eqn [176a], because that would reduce the order of the ODE and make it impossible to satisfy all the boundary conditions. Equations [176a] and [176b] therefore constitute a singular perturbation problem: the solution for small values of $\epsilon > 0$ is not a small perturbation of a solution for $\epsilon = 0$, which in any case does not exist. The resolution of this apparent paradox is that the solution exhibits a thin boundary layer where the term $\epsilon T''$ in eqn [176a] is important, no matter how small ϵ may be.

We therefore anticipate that the solution to eqn [176] will comprise two distinct regions governed by different dynamics: an inner region (the BL) of dimensionless thickness $\delta \ll 1$ in which advection is balanced by diffusion, and an outer region where advection is balanced by adiabatic decompression. Consider the outer region first, and let the temperature there be $T = h(z,\epsilon)$. We seek a solution in the form of an asymptotic expansion

$$h = h_0(z) + \eta_1(\epsilon)h_1(z) + \eta_2(\epsilon)h_2(z) + \cdots \qquad [178]$$

where $\eta_n(\epsilon)$ are (as yet unknown) gauge functions that form an asymptotic sequence such that $\lim_{\epsilon\to 0}(\eta_n/\eta_{n-1})=0$. The function $\eta_0(\epsilon)=1$ because $T=O(1)$ in the outer region. Substituting eqn [178] into eqn [176a] and retaining only the lowest-order terms, we obtain

$$h_0' - \beta(h_0+\lambda)=0 \qquad [179]$$

Because eqn [179] is a first-order ODE, its solution can satisfy only one of the boundary conditions [176b]. Evidently, this must be the condition at $z=1$, because $z=0$ is the upper limit of a TBL that cannot be described by eqn [179]. The same conclusion can be reached in a more formal algorithmic way by assuming contrary to fact that the BL is at $z=1$, solving eqn [179] subject to the wrong boundary condition $T(0)=0$, and finally realizing that the inner solution for the supposed BL is unphysical because it increases exponentially upward. The lowest-order terms of the (correct) boundary condition are $h_0(1)=1$, and the solution of eqn [179] that satisfies this is

$$h_0=(1+\lambda)\exp\beta(z-1)-\lambda \qquad [180]$$

Turning now to the inner region, the first task is to determine which term on the LHS of eqn [176a] balances the RHS. Suppose (contrary to fact, as will soon appear) that $\beta(T+\lambda)\gg T'$ (decompression $\gg$ advection). Because $z\sim\delta$ and $z^2\gg z^3$ in the BL, the balance $z^2\beta(T+\lambda)\sim\epsilon T''$ implies $\delta\sim\epsilon^{1/4}$. But then $T'\gg\beta(T+\lambda)$, which contradicts our original assumption. The correct balance must therefore be $z^2T'\sim\epsilon T''$ (advection $\sim$ diffusion), which implies $\delta\sim\epsilon^{1/3}$ and $T'\gg\beta(T+\lambda)$, consistent with our original assumption. A more physical argument that leads to the same conclusion is to note that the vertical scale length over which adiabatic decompression is significant in the mantle ($\sim$1000 km) is much greater than a typical BL thickness.

Now that we know the thickness of the BL, we can proceed to determine its structure. To discern the thin BL distinctly, we use a sort of mathematical magnifying glass: a new stretched depth coordinate $\hat{z}=z/\delta\equiv z\epsilon^{-1/3}$ that is of order unity in the BL. Denoting the inner solution by $f(\hat{z},\epsilon)$ and writing eqn [176a] in terms of $\hat{z}$, we find

$$\left(2\epsilon\hat{z}^3-3\epsilon^{2/3}\hat{z}^2\right)\left[\epsilon^{-1/3}f'-\beta(f+\lambda)\right]=\epsilon^{1/3}f'' \qquad [181]$$

where primes denote differentiation with respect to $\hat{z}$. We now seek a solution in the form

$$f=f_0(z)+v_1(\epsilon)f_1(z)+v_2(\epsilon)f_2(z)+\cdots \qquad [182]$$

where $v_n(\epsilon)$ are unknown gauge functions. Substituting eqn [182] into eqn [181] and retaining only the lowest-order terms, we obtain

$$f_0''+3\hat{z}^2f_0'=0 \qquad [183]$$

The solution of eqn [183] that satisfies the boundary condition $f_0(0)=0$ is

$$f_0=A\int_0^{\hat{z}}\exp\left(-x^3\right)\mathrm{d}x\equiv\frac{A}{3}\gamma\left(1/3,\hat{z}^3\right) \qquad [184]$$

where $\gamma(a,x)=\int_0^x\exp(-t)t^{a-1}\mathrm{d}t$ is the incomplete gamma function and A is an unknown constant that must be determined by matching eqn [184] to the outer solution [180].

The most rigorous way to do the matching is to rewrite both the inner and outer expansions in terms of an intermediate variable $z_{\mathrm{int}}=\xi(\epsilon)z$ such that $\epsilon^{-1/3}\ll\xi\ll 1$ and then to choose the values of any unknown constants (A in this case) so that the two expressions agree. However, it is often possible to use a simpler matching principle, due to Prandtl, which states that the inner limit of the outer expansion must be equal to the outer limit of the inner expansion – roughly speaking, that the two expansions must match at the edge of the BL. For our problem, Prandtl's principle is

$$\lim_{\hat{z}\to\infty}f_0(\hat{z})=\lim_{z\to 0}h_0(z),\quad \text{whence } A=\frac{3[(1+\lambda)\exp(-\beta)-\lambda]}{\Gamma(1/3)} \qquad [185]$$

where Γ is the gamma function.

The last step is to construct a composite expansion that is valid both inside and outside the BL. This is just the sum of the inner and outer expansions less their shared common part $h_0(z\to 0)\equiv(1+\lambda)\exp(-\beta)-\lambda$, or

$$T=h_0(z)+[(1+\lambda)\exp(-\beta)-\lambda]\left[\frac{\gamma(1/3,z^3/\epsilon)}{\Gamma(1/3)}-1\right] \qquad [186]$$

The procedure just described is a first-order matching that retains only the first terms in the expansions [178] and [182]. If desired, the matching can be carried out to higher order by working back and forth between the inner and outer expansions, determining the gauge functions $\eta_n(\epsilon)$ and $v_n(\epsilon)$ and matching at each step. Higher-order matching often requires the use of the more rigorous intermediate matching principle; for examples, see Hinch (1991) and Kevorkian and Cole (1996), or the somewhat less formal treatments of Nayfeh (1973) or Van Dyke (1975). For many problems, however, first-order matching suffices to reveal the essential structure and physical significance of the solution.

The MMAE has been used to solve a variety of geophysically relevant problems involving BLs, mostly with constant viscosity. Umemura and Busse (1989) studied the axisymmetric convective flow in a cylindrical container of height d with free-slip boundaries, using the MMAE to match the interior flow to the central rising plume and the circumferential downwelling. They found that the vertical velocity in the plume is $w\sim(\kappa/d)(-\ln\epsilon)^{1/2}\mathrm{Ra}^{2/3}$, where the dimensionless plume radius $\epsilon\equiv a/d$ satisfies $\epsilon(-\ln\epsilon)^{1/4}\sim\mathrm{Ra}^{-1/6}$, $\mathrm{Ra}\gg 1$ being the Rayleigh number. Whittaker and Lister (2006a) studied creeping plumes from a point heat source with buoyancy flux B on an impermeable plane boundary by matching the BL flow to an outer flow that sees the plume as a line distribution of Stokeslets. They found that the vertical velocity $w\sim(B/\nu)^{1/2}[\ln(z/z_0)]^{1/2}$, where z is the height and $z_0=32\pi\kappa^2\nu/B$. Whittaker and Lister (2006b) studied the dynamics of a plume above a heated disk on a plane boundary and used the MMAE to match the flow within the plume to both an outer flow and a horizontal BL flow across the disk. The results show that the Nusselt number $\mathrm{Nu}\sim\mathrm{Ra}^{1/5}$ for a rigid boundary and $\mathrm{Nu}\sim(\mathrm{Ra}/\ln\mathrm{Ra})^{1/3}$ for a free-slip boundary, where $\mathrm{Ra}\gg 1$ is defined using the disk radius. An application of the MMAE to a problem with variable viscosity (Morris, 1982) is given in the next subsection.

7.04.7.4 Boundary Layers with Strongly Variable Viscosity

In a fluid with constant viscosity and infinite Prandtl number, thermal BLs are not accompanied by vorticity BLs, because the velocity field varies on a length scale much larger than the

thickness of the TBL. As a result, the velocity field within the BL can be represented by the first term of its Taylor series expansion (or its multipole expansion in the case of an axisymmetric plume; cf. Whittaker and Lister, 2006a), which greatly simplifies the task of solving the thermal BL equation. This happy state of affairs no longer obtains if the viscosity of the fluid depends on temperature, because the velocity and temperature fields are then coupled.

Most studies of variable-viscosity BLs in the geodynamics literature focus on mantle plumes in either planar (**Figure 11(a)**) or axisymmetric (**Figure 11(b)**) geometry. The basic procedure is to supplement the BL eqns [158] and [159] with a simplified BL form of the vertical component of the momentum equation in which derivatives along the BL are neglected relative to those normal to it. In physical terms, this equation simply states that the buoyancy force in the plume is balanced by the lateral gradient of the vertical shear stress τ or

$$y^{-m}\partial_y(y^m\tau)+\rho g\alpha(T-T_\infty)=0 \qquad [187]$$

where T_∞ is the temperature far from the plume, y is the coordinate normal to the plume, and $m=0$ or 1 for planar or axisymmetric geometries, respectively.

An important early study based on eqn [187] in planar geometry was that of Yuen and Schubert (1976), who investigated the buoyant upwelling of a fluid with temperature-dependent Newtonian or power-law rheology adjacent to an isothermal free-slip boundary. The BL equations for this case admit a similarity transformation of the form

$$T=T_\infty+f(\eta),\quad \psi=x^{(n+2)/(n+3)}g(\eta),\quad \eta=yx^{-1/(n+3)} \qquad [188]$$

where ψ is the stream function and n is the power-law exponent in the constitutive law [43f].

The more realistic case of an axisymmetric plume from a point source of heat in a fluid with temperature-dependent viscosity has been studied by several authors. Using a similarity transformation, Morris (1980) found that the temperature on the plume axis decreases exponentially upward with a scale height $Q\Delta T_r/12\pi k_c$, where Q is the total heat flux carried by the plume, ΔT_r is the temperature change required to change the viscosity by a factor e, and k_c is the thermal conductivity. Crosby and Lister (2014) examined the same problem in more detail, confirming Morris's (1980) result and finding good agreement between numerical solutions and an analytical similarity solution of the boundary-layer equations. Hauri et al. (1994) considered a similar problem, but with an empirical superexponential temperature- and depth-dependent viscosity law.

While plumes involve free or buoyancy-driven convection, variable-viscosity boundary layers can also arise in situations where a large-scale background flow or wind is imposed externally (forced convection). An example is the previously introduced stagnation-point flow model for the heat transfer from a hot sphere moving in a fluid with strongly temperature-dependent viscosity (**Figure 1(b)**). Using the MMAE, Morris (1982) showed that three distinct dynamical regimes occur: a conduction limit; a Stokes limit in which the flow around the sphere resembles that in an isoviscous fluid, and a lubrication limit in which most of the volume flux is carried by the low-viscosity BL adjoining the hot plate. Solutions in spherical geometry were obtained by Morris (1982) and Ansari and Morris (1985) for the lubrication limit and an intermediate (lubrication/Stokes) limit.

7.04.8 Long-Wave Theories

Long-wave theories comprise a variety of loosely related approaches whose goal is to describe the evolving thickness or shape of a thin layer. The fundamental idea underlying these approaches is the distinction between small amplitude and small slope. To take a simple example, if a 2-D fluid layer has thickness $h_0+\Delta h\sin kx$, its surface has small amplitude if $\Delta h/h_0\ll 1$ and small slope (or long wavelength) if $k\Delta h\ll 1$. Long-wave theories typically exploit the fact that small slope does not imply small amplitude to derive equations governing the nonlinear (finite-amplitude) evolution of long-wave disturbances in a layer.

7.04.8.1 Lubrication Theory

Shear-dominated viscous flows in thin layers are common in geodynamics: examples include lava flows and the spreading of buoyant plume material beneath the lithosphere. Such flows are described by a simplified form of the Navier–Stokes equations, called the lubrication equations because of their importance in the design of industrial lubrication bearings. To develop the theory, we return to the simple model of a viscous drop or gravity current spreading on a rigid surface sketched in Figure 3(b). The lubrication equations that describe such a current can be obtained from the full Navier–Stokes equations by exploiting the fact that $h/R\equiv\epsilon\ll 1$, where h and R are the layer's characteristic thickness and lateral extent, respectively. The thinness of the layer has three important consequences: (1) The horizontal fluid velocity $\mathbf{u}\equiv u\mathbf{e}_x+v\mathbf{e}_y$ along the layer greatly exceeds the velocity w normal to it. The flow is therefore nearly parallel to the surface, and inertia is negligible. (2) Derivatives of the velocity components across the layer greatly exceed the derivatives along the layer ($\partial_z\sim h^{-1}\gg\partial_x\sim\partial_y\sim R^{-1}$). (3) The pressure gradient across the layer is approximately hydrostatic. The Navier–Stokes equations then take the simplified forms

$$\nabla_h\cdot\mathbf{u}+w_z=0,\quad \nabla_h p=\eta\mathbf{u}_{zz},\quad p_z=-\rho g \qquad [189]$$

where w is the vertical velocity.

To illustrate the use of the lubrication equations, we determine the shape and spreading rate of an axisymmetric gravity current with constant volume V (Huppert, 1982). The geometry of this situation is that of **Figure 3(b)** except that the vertical conduit supplying fluid to the current is absent. Let r be the (horizontal) radial coordinate, $u(z,r,t)$ the radial component of the velocity, $h(r,t)$ the thickness of the current, and $R(t)$ its radius. In this (cylindrical) geometry, eqn [189] take the forms

$$r^{-1}(ru)_r+w_z=0,\quad p_r=\eta u_{zz},\quad p_z=-\rho g \qquad [190]$$

Integrating eqn [190] subject to $p(z=h)=0$, we obtain

$$p=\rho g(h-z). \qquad [191]$$

Next, we substitute eqn [191] into [190] and integrate subject to the no-slip condition on the plate ($u(0,r,t)=0$) and vanishing traction at $z=h$. Now, because the current's upper surface is nearly horizontal, the traction there

$\approx\eta(u_z+w_x)$. However, $w_x\sim(h/R)^2u_z$ in the lubrication approximation, so the condition of vanishing traction is simply $u_z(h,r,t)=0$. The profile of radial velocity across the current is therefore

$$u=\frac{\rho g h_r}{2\eta}\left(z^2-2hz\right) \qquad [192]$$

where $\rho g h_r$ is the radial gradient of the hydrostatic pressure that drives the flow. Next, the continuity eqn [190] is integrated across the current subject to the impermeability condition $w(0,r,t)=0$ to obtain

$$0=w(h,r,t)+r^{-1}\int_0^h (ru)_r\,\mathrm{d}z \qquad [193]$$

We now simplify eqn [193] by using the kinematic surface condition to eliminate $w(h,r,t)\equiv h_t+u(h,r,t)h_r$, taking ∂_r outside the integral using the standard formula for the derivative of an integral with a variable limit, and evaluating the integral using eqn [192]. We thereby find that $h(r,t)$ satisfies the nonlinear diffusion equation

$$h_t=4\sigma r^{-1}\left(rh^3h_r\right)_r,\quad \sigma=\frac{\rho g}{12\eta} \qquad [194]$$

where σ is the spreadability. Conservation of the current's volume requires

$$2\pi\int_0^R rh\,\mathrm{d}r=V. \qquad [195]$$

We anticipate that for long times, the current will achieve a universal self-similar shape that retains no memory of the initial shape $h(r,0)$. Self-similarity requires that h depend on the normalized radius $\zeta\equiv r/R(t)\in[0,1]$, and conservation of volume requires $hR^2\sim V$. We therefore seek a solution of the form

$$h=\frac{V}{R^2}H\left(\frac{r}{R}\right). \qquad [196]$$

Upon substituting eqn [196] into eqns [194] and [195], separating variables in the now-familiar way (see **Section 7.04.4.3**), and solving the resulting equations subject to the conditions $H'(0)=H(1)=R(0)=0$, we obtain

$$H=\frac{4}{3\pi}\left(1-\zeta^2\right)^{1/3},\quad R=\left(\frac{4096\sigma V^3t}{81\pi^3}\right)^{1/8} \qquad [197]$$

The radius of the gravity current increases as the 1/8 power of the time.

Solution [197] is a special case of a more general class of similarity solutions of eqn [194] studied by Gratton and Minotti (1990) using a phase-plane formalism. Rather than work with the single PDE [194], Gratton and Minotti (1990) wrote down two coupled PDEs for the thickness h and the mean horizontal velocity v and then used a similarity transformation to reduce them to coupled ODEs. The solutions of these ODEs can be represented as segments of integral curves on a phase plane that connect singular points representing different boundary conditions such as sources, sinks, and current fronts. Gratton and Minotti (1990) gave an exhaustive catalog of the solutions thus found, including a novel second-kind similarity solution (cf. **Section 7.04.4.3**) for the evolution of an axisymmetric gravity current surrounding a circular hole.

Geophysical applications of viscous gravity current theory include the solutions of Lister and Kerr (1989b) for the spreading of 2-D and axisymmetric currents at a fluid interface, which were applied by Kerr and Lister (1987) to the spreading of subducted lithosphere along the boundary between the upper and lower mantle. In the next subsection, we discuss an extension of the theory to currents spreading on moving surfaces, which have been widely used to model the interaction of mantle plumes with a moving or rifting lithosphere.

7.04.8.2 Plume–Plate and PRI Models

The geometry of these models was introduced in **Section 7.04.2.3**, and is sketched in **Figure 3(c)** and **3(d)**. The motion of a plate (or plates) with a horizontal velocity $\mathbf{U}_0(x,y)$ generates an ambient flow $\mathbf{U}(x,y,z)$ in the mantle below, which is assumed to have uniform density ρ and viscosity η_m. The plume conduit is represented as a volume source of strength Q fixed at $(x,y)=(0,0)$ that emits buoyant fluid with density $\rho-\Delta\rho$ and viscosity η_p. This fluid spreads laterally beneath the lithosphere to form a thin layer whose thickness $h(x,y,t)$ is governed by a balance of buoyancy-driven spreading and advection by the ambient mantle flow. Strictly speaking, the solution of this problem requires the simultaneous determination of the flow in both fluids subject to the usual matching conditions on velocity and stress at their interface. However, useful results can be obtained via a simpler approach in which the mantle flow $\mathbf{U}(x,y,z)$ is specified a priori and is assumed to be unaffected by the flow in the plume layer. Olson (1990) proposed a model of this type for a plume beneath a plate moving at constant speed U_0 (**Figure 3(c)**) and derived a lubrication equation for $h(x,y,t)$ assuming a uniform mantle flow $\mathbf{U}(x,y,z)=U_0\mathbf{e}_x$. For the important special case of a steady-state plume layer, the lubrication equation is

$$U_0h_x=\overline{\sigma}\nabla_{\mathrm{h}}^2h^4+Q\delta(x)\delta(y),\quad \overline{\sigma}=\frac{g\Delta\rho}{48\eta_p} \qquad [198]$$

The three terms in eqn [198] represent advection, buoyancy-driven lateral spreading, and injection of the plume fluid, respectively. The spreadability $\overline{\sigma}$ is four times smaller than that for the gravity current (eqn [194]) because Olson (1990) applied a no-slip boundary condition $\mathbf{u}(x,y,h)=U_0\mathbf{e}_x$ at $z=h$, where the plume fluid is in contact with the much more viscous mantle.

The fundamental scales for the thickness h and width L of the plume fluid layer can be found via a scaling analysis of eqn [198]. By requiring the three terms in eqn [198] to be of the same order and noting that $\partial_x\sim\partial_y\sim\delta(x)\sim\delta(y)\sim L^{-1}$, one finds (Ribe and Christensen, 1994)

$$L\sim\frac{Q^{3/4}\overline{\sigma}^{-1/4}}{U_0}\equiv L_0,\quad h\sim\left(\frac{Q}{\overline{\sigma}}\right)^{1/4}\equiv h_0 \qquad [199]$$

Further insight is provided by an analytical similarity solution of eqn [198] that is valid far downstream ($x\gg L_0$) from the plume source. We anticipate that at these distances, the layer thickness h will vary more strongly in the direction normal to

the plate motion than parallel to it, so that $\nabla_h^2 h^4 \sim (h^4)_{yy}$. Equation [11] now reduces to

$$U_0 h_x = \left(h^4\right)_{yy}, \quad U_0 \int_{-L/2}^{L/2} h\,\mathrm{d}y = Q \qquad [199]$$

where the source strength Q now appears in an integral relation expressing conservation of the downstream volume flux. Equation [200] admits the similarity solution (Ribe and Christensen, 1994):

$$\hat{h} = \left(\frac{3}{40}\right)^{1/3} \hat{x}^{-1/5}\left(C^2 - \hat{y}^2 \hat{x}^{-2/5}\right)^{1/3},$$
$$C = \left(\frac{390625}{9\pi^3}\right)^{1/10}\left[\frac{\Gamma(5/6)}{\Gamma(1/3)}\right]^{3/5} \qquad [200]$$

where $\hat{h} = h/h_0$ and $(\hat{x}, \hat{y}) = (x, y)/L_0$. The width of the plume layer increases with downstream distance as $L = 2CL_0(x/L_0)^{1/5}$. Asaadi et al. (2011) showed that the rate of increase is slower ($\sim x^{1/(3n+2)}$) if the plume has a power-law rheology [43f] and compared numerical solutions of the power-law analogue of eqn [200] with the shape of the Hawaiian swell to conclude that the material compensating the swell is non-Newtonian. The previously mentioned scaling laws for $L(x)$ remain valid for more realistic 3-D numerical models with moderately temperature- and pressure-dependent viscosity (Asaadi et al., 2011; Ribe and Christensen, 1994, 1999) but will break down if the mantle/plume viscosity ratio significantly exceeds the aspect ratio (width/thickness) of the plume layer.

The interaction of a plume with an oceanic spreading ridge has been studied using a modified version of Olson's (1990) model in which the uniform mantle flow $\mathbf{U} = U_0\mathbf{e}_x$ is replaced by a corner flow driven by surface plates diverging with a half-spreading rate U_0 (**Figure 3(d)**; Ribe et al., 1995). The balance of the source term $Q\delta(x)\delta(y)$ with vertical advection directly beneath the ridge yields a new length scale $L_1 \sim (Q/U_0)^{1/2}$, which agrees well with the along-strike extent of the plume fluid beneath the ridge (waist width) determined by laboratory experiments (Feighner and Richards, 1995) and numerical models (Ribe et al., 1995).

An important extension of the theory for geodynamical applications is to spreading currents in which both the buoyancy and viscosity depend on temperature. Such a theory was developed by Bercovici (1994) and Bercovici and Lin (1996), who supplemented the usual lubrication theory equations with an energy equation describing the temperature distribution inside the current. They found that variable viscosity and buoyancy strongly influence the current's shape and spreading rate, which no longer exhibit the self-similar behavior typical of currents with constant properties.

7.04.8.3 Long-Wave Analysis of Buoyant Instability

In convecting systems such as the Earth's mantle, plumes arise as instabilities of horizontal TBLs. A long-wave model for this process has been proposed by Lemery et al. (2000; henceforth LRS00), based on two assumptions: that the characteristic horizontal scale of the flow greatly exceeds the BL thickness, and that the horizontal velocity of the fluid is approximately constant across the BL. The latter assumption excludes hot BLs within which the (temperature-dependent) viscosity is much lower than outside, and makes the model most applicable to plume formation in cold BLs. The two assumptions allow the coupled 3-D dynamics of the BL and the fluid outside it to be reduced to 2-D equations for the lateral velocity at the edge of the BL and a temperature moment that describes the distribution of buoyancy within it.

The domain of the model is a fluid half-space bounded by a cold free-slip surface $x_3 = 0$ held at temperature $-\Delta T$ relative to the fluid far from it (**Figure 2(b)**). The BL occupies the depth interval $x_3 < h(x_1, x_2, t)$, where x_α are Cartesian coordinates parallel to the BL and t is time. In the following, hatted and unhatted variables are those in the BL and in the interior, respectively, and an argument in parentheses indicates a value of x_3. The (constant) viscosity of the outer fluid is η_0, and the viscosity within the BL is $\hat{\eta}(x_3)$.

The starting point is the momentum equation within the BL, namely, $-\partial_i \hat{p} + \partial_j \hat{\tau}_{ij} = \rho_0 \alpha g \hat{T} \delta_{i3}$, where $\hat{\tau}_{ij}$ is the deviatoric part of the stress tensor. Taking the curl of this equation, applying the continuity equation $\hat{\tau}_{33} = -\hat{\tau}_{\gamma\gamma}$, and noting that $\left(\partial_{11}^2 \hat{\tau}_{\alpha 3}, \partial_{12}^2 \hat{\tau}_{\alpha 3}, \partial_{22}^2 \hat{\tau}_{\alpha 3}\right) \ll \partial_{33} \hat{\tau}_{\alpha 3}$ in the long-wavelength approximation, we obtain

$$\partial_{33}^2 \hat{\tau}_{\alpha 3} + \partial_3 \hat{A}_\alpha = -\rho_0 \alpha g \partial_\alpha \hat{T}, \qquad \hat{A}_\alpha = \partial_\alpha \hat{\tau}_{\gamma\gamma} + \partial_\gamma \hat{\tau}_{\alpha\gamma} \qquad [202]$$

Physically, eqn [202] are the lateral ($\alpha = 1$ or 2) components of the vorticity equation. Now, multiply eqn [202] by x_3, integrate across the boundary layer from $x_3 = 0$ to $x_3 = h$, and take lateral derivatives outside the integral signs by neglecting the small lateral variation of the upper limit $h(x_1, x_2, t)$. The result is

$$-\hat{\tau}_{\alpha 3}(h) + h\partial_3 \hat{\tau}_{\alpha 3}(h) + h\hat{A}_\alpha(h) - \left\langle \hat{A}_\alpha \right\rangle = \rho_0 \alpha g \partial_\alpha M \qquad [203]$$

where $\langle\ \rangle = \int_0^h \mathrm{d}x_3$ and

$$M = -\left\langle x_3 \hat{T} \right\rangle \qquad [204]$$

is the temperature moment. Now, $h\partial_3 \hat{\tau}_{\alpha 3}(h) \ll \hat{\tau}_{\alpha 3}(h)$ in the long-wave limit, and continuity of shear stress at $x_3 = h$ requires $\hat{\tau}_{\alpha 3}(h) = \tau_{\alpha 3}(h)$. But because the interior fluid sees the BL as a skin with zero thickness, $\tau_{\alpha 3}(h) \approx \tau_{\alpha 3}(0) \approx \eta_0 \partial_3 u_\alpha(0)$. Moreover, the lateral velocity components are constant across the layer to lowest order and must match those in the interior fluid, requiring $\hat{u}_\alpha = u_\alpha(0)$ and $\hat{\tau}_{\alpha\beta} = \hat{\eta}(x_3)\left[\partial_\alpha u_\beta(0) + \partial_\beta u_\alpha(0)\right]$. Substituting these expressions into eqn [203], we find

$$-\partial_3 \mathbf{u} - \nabla_h \cdot (\sigma \mathbf{T}) - 3\nabla_h(\sigma \nabla_h \cdot \mathbf{u}) = \frac{\rho_0 g \alpha}{\eta_0} \nabla_h M \qquad [205]$$

where

$$\sigma = \frac{\hat{\eta}(x_3) - \eta_0}{\eta_0} \equiv \frac{1}{\eta_0} \int_0^h (\hat{\eta}(x_3) - \eta_0)\mathrm{d}x_3 \qquad [206]$$

is a quantity with units of length that measures the excess viscosity of the boundary layer, $\mathbf{u}$ is the horizontal velocity vector, $\mathbf{T}$ is a modified strain rate tensor with components $T_{11} = -T_{22} = \partial_1 u_1 - \partial_2 u_2$ and $T_{12} = T_{21} = \partial_1 u_2 + \partial_2 u_1$, and all terms are evaluated at $x_3 = 0$. Equation [205] is an effective boundary condition that represents the influence of the BL on the interior fluid. It shows that the BL acts like an extensible skin with (depth-integrated) shear viscosity $\sigma\eta_0$ and compressional viscosity $3\sigma\eta_0$ that applies a shear stress proportional to $\nabla_h M$ to the outer fluid.

The next step is to determine explicitly the flow in the outer fluid that is driven by lateral variations in M by solving the Stokes equations in the halfspace $x_3 \geq 0$ subject to eqn [205]. Evaluating the resulting solution at $x_3 = 0$, we obtain the closure relationship (LRS00, eqn 2.41)

$$\overline{\mathbf{u}}(\mathbf{k}) = \frac{i\mathbf{k}}{2k(1+2\sigma k)} \frac{\rho_0 g \alpha}{\eta_0} \overline{M}(\mathbf{k}) \qquad [207]$$

where $\mathbf{k}$ is the horizontal wavevector, $k = |\mathbf{k}|$, and overbars denote the Fourier transform. Equation [207] is only valid if σ does not vary laterally. Finally, an evolution equation for M is obtained by taking the first moment of the energy equation, yielding

$$\partial_t M + \mathbf{u}\cdot\nabla_h M + 2M\nabla_h\cdot\mathbf{u} = \kappa\left(\nabla_h^2 M + \Delta T\right) \qquad [208]$$

eqns [207] and [208] are three equations for $u_\alpha(x_1, x_2, t)$ and $M(x_1, x_2, t)$ that can be solved numerically subject to periodic boundary conditions in the lateral directions for a specified initial condition $M(x_1, x_2, 0)$.

An important special case of the equations just derived is the R–T instability of a layer with density $\rho_0 + \Delta\rho$ and viscosity $\eta_1 = \gamma\eta_0$, obtained by the transformation

$$M \to h^2 \Delta\rho / 2\rho_0 \alpha, \quad \sigma \to h(\gamma - 1), \quad \kappa \to 0 \qquad [209]$$

A linear stability analysis of eqns [207] and [208] can now be performed by setting $h = h_0 + \tilde{h}\exp(i\mathbf{k}\cdot\mathbf{x})\exp(st)$ and $\mathbf{u} = \overline{\mathbf{u}}\exp(i\mathbf{k}\cdot\mathbf{x})\exp(st)$ and linearizing in the perturbations $\tilde{h}$ and $\tilde{\mathbf{u}}$. The resulting growth rate is $s/s_1 = \epsilon\gamma/2[1 + 2\epsilon(\gamma - 1)]$ where $\epsilon = h_0 k$ and $s_1 = g\Delta\rho h_0/\eta_1$, which agrees with the exact analytical expression [39] if $\gamma \gg \epsilon$. Equations [205]–[208] are therefore valid as long as the BL is not too much less viscous than the outer fluid.

The equations just derived also describe the finite-amplitude evolution of the R–T instability of a dense viscous layer over a passive half-space (Canright and Morris, 1993 ≡ CM93). Because the half-space is effectively inviscid ($\gamma \to \infty$), the closure law [207] is not meaningful, and the relevant equations are eqns [205] and [208]. Rewriting these using eqn [209] and noting that the first term in eqn [205] (shear stress applied by the inviscid fluid) is negligible, we obtain

$$\nabla_h\left(\frac{g\Delta\rho}{4\eta_1}h^2 + h\nabla_h\cdot\mathbf{u}\right) + \nabla_h\cdot(h\mathbf{E}) = 0 \qquad [210]$$

$$\partial_t h + \nabla_h\cdot(h\mathbf{u}) = 0 \qquad [211]$$

where $\mathbf{E}$ is the standard 2D strain rate tensor. Equations [210] and [211] are just the dimensional forms of eqns. (3.8) and (3.7), respectively, of CM93.

A remarkable feature of eqns [207], [208] and [210], [211] is the existence of solutions in which M or h becomes infinite at a finite time t_*, corresponding to the runaway escape of the plume from its source layer. Even though the flow in the vicinity of the developing singularity violates the original long-wave assumption of the model, this failure is only local and does not impair the model's global validity. Several examples are discussed by CM93, who obtained solutions of eqns [210] and [211] which show that the amplitude of disturbances of a layer with initial thickness h_0 evolves at long times as

$$h \sim \frac{1}{t_* - t} \qquad [212]$$

where $t_* = C\eta_1/h_0 g\Delta\rho$ and the value of C depends on the form of the initial disturbance.

7.04.8.4 Theory of Thin Shells, Plates, and Sheets

A central problem in geodynamics is to determine the response of the lithosphere to applied loads such as seamounts, plate boundary forces (ridge push, slab pull, etc.), and tractions imposed by the underlying mantle convection. Such problems can be solved effectively using thin-shell theory, a branch of applied mechanics concerned with the behavior of freely deformable bodies whose thickness h is much smaller than both their typical lateral dimension L and their typical radius of curvature R. This condition is evidently satisfied for the Earth's lithosphere, for which $h \approx 100$ km, $L \approx 1000$–$10\,000$ km, and $R \approx 6300$ km. More generally, thin-shell theory applies to layers in which the stress is dominated by the components $\sigma_{\alpha\beta}$, where α and β are coordinate directions parallel to the layer. This is the case when the layer is not attached to a stiff substrate and is free to deform by a combination of stretching and bending. By contrast, lubrication theory is valid when the stress is dominated by the layer-parallel shear components $\sigma_{\alpha 3}$, as occurs when the layer adjoins a substrate or wall on which the no-slip condition applies.

The basic idea of thin-shell theory is to exploit the smallness of h/L and h/R to reduce the full 3-D dynamical equations to equivalent 2-D equations for the dynamics of the shell's midsurface. Let the 3-D Cartesian coordinates of any point on this surface be $\mathbf{x}_0(\theta_1, \theta_2)$, where θ_α are coordinates on the midsurface itself. In the most general formulations of shell theory (e.g., Niordson, 1985), θ_α are allowed to be arbitrary and nonorthogonal. Such a formulation is useful for problems involving large finite deformation, because θ_α can be treated as Lagrangian coordinates. If the deformation is small, however, it makes sense to define θ_α as orthogonal lines-of-curvature coordinates whose isolines are parallel to the two directions of principal curvature of the midsurface at each point. This less elegant but more readily understandable formulation is the one used in most geodynamical applications of shell theory.

Relative to lines-of-curvature coordinates, the fundamental quantities that describe the shape of the midsurface are the principal radii of curvature R_α and the Lamé parameters $A_\alpha = |\partial_\alpha \mathbf{x}_0|$, where $\partial_\alpha = \partial/\partial_{\theta_\alpha}$. For notational convenience, let $B_\alpha = 1/A_\alpha$ and $K_\alpha = 1/R_\alpha$. All vector and tensor quantities defined on the midsurface are expressed relative to a local orthonormal basis comprising two surface-parallel unit vectors $\mathbf{d}_1, \mathbf{d}_2$ and a normal vector $\mathbf{d}_3$ defined by

$$\mathbf{d}_1 = B_1\partial_1\mathbf{x}_0, \quad \mathbf{d}_2 = B_2\partial_2\mathbf{x}_0, \quad \mathbf{d}_3 = \mathbf{d}_1 \times \mathbf{d}_2 \qquad [213]$$

In the following, z is a coordinate normal to the midsurface $z = 0$. Moreover, the repeated subscript α is not summed unless explicitly indicated, and $\beta = 2$ when $\alpha = 1$ and vice versa.

The general equilibrium equations for a shell without inertia are (Novozhilov, 1959, p. 39)

$$\begin{aligned}&\partial_\alpha(A_\beta T_\alpha) + \partial_\beta(A_\alpha S) + S\partial_\beta A_\alpha - T_\beta\partial_\alpha A_\beta \\ &+ K_\alpha\left[\partial_\alpha(A_\beta M_\alpha) - M_\beta\partial_\alpha A_\beta + 2\partial_\beta(A_\alpha H) + 2R_\alpha K_\beta H\partial_\beta A_\alpha\right] \\ &= -A_1 A_2 P_\alpha \quad (\alpha = 1 \text{ or } 2)\end{aligned} \qquad [214a]$$

$$\sum_{\alpha=1}^{2}\{B_1B_2\partial_\alpha[B_\alpha(\partial_\alpha(A_\beta M_\alpha)-M_\beta\partial_\alpha A_\beta+\partial_\beta(A_\alpha H)+H\partial_\beta A_\alpha)] -K_\alpha T_\alpha\}=-P_3 \quad [214b]$$

where

$$T_\alpha=\int_{-h/2}^{h/2}(1+K_\beta z)\sigma_{\alpha\alpha}\mathrm{d}z,$$
$$S=\int_{-h/2}^{h/2}(1-K_1K_2z^2)\sigma_{12}\mathrm{d}z \quad [215a]$$
$$M_\alpha=\int_{-h/2}^{h/2}z(1+K_\beta z)\sigma_{\alpha\alpha}\mathrm{d}z,$$
$$H=\int_{-h/2}^{h/2}z[1+(K_1+K_2)z/2]\sigma_{12}\mathrm{d}z \quad [215b]$$

P_i is the total load vector (per unit midsurface area), and $\sigma_{\gamma\lambda}$ is the usual Cauchy stress tensor. The essential content of eqn [214] is that a loaded shell can deform in two distinct ways: by in-plane stretching and shear, the intensity of which is measured by the stress resultants T_α and S, and by bending, which is measured by the bending moments M_α and H. In general, both modes are present, in a proportion that depends in a complicated way on the midsurface shape and on the distribution of the applied loads. Shells that deform only by extension and in-plane shear ($M_\alpha=H=0$) are called membranes.

Equations [214a] and [214b] are valid for a shell of any material. To solve them, we need constitutive relations that link T_α, S, M_α, and H to the displacement ζ_i (for an elastic shell) or the velocity u_i (for a fluid shell) of the midsurface. For an elastic shell with Young's modulus E and Poisson's ratio ν, these are (Novozhilov, 1959, p. 24, 48)

$$T_\alpha=\frac{Eh}{1-\nu^2}(\epsilon_\alpha+\nu\epsilon_\beta),\quad M_\alpha=\frac{Eh^3}{12(1-\nu^2)}(\kappa_\alpha+\nu\kappa_\beta) \quad [216a]$$

$$S=\frac{Eh}{2(1+\nu)}\omega,\quad H=\frac{Eh^3}{12(1+\nu)}\tau \quad [216b]$$

where

$$\epsilon_\alpha=B_\alpha\partial_\alpha\zeta_\alpha+B_1B_2\zeta_\beta\partial_\beta A_\alpha+K_\alpha\zeta_3 \quad [217a]$$

$$\omega=\sum_{\alpha=1}^{2}A_\beta B_\alpha\partial_\alpha(B_\beta\zeta_\beta) \quad [217b]$$

$$\kappa_\alpha=-B_\alpha\partial_\alpha(B_\alpha\partial_\alpha\zeta_3-K_\alpha\zeta_\alpha)-B_1B_2\partial_\beta A_\alpha(B_\beta\partial_\beta\zeta_3-K_\beta\zeta_\beta) \quad [217c]$$

$$\tau=-B_1B_2\partial_{12}^2\zeta_3+\sum_{\alpha=1}^{2}\{B_1B_2B_\alpha(\partial_\beta A_\alpha)\partial_\alpha\zeta_3 +K_\alpha(B_\beta\partial_\beta\zeta_\alpha-B_1B_2\zeta_3\partial_\beta A_\alpha)\} \quad [217d]$$

are the six independent quantities that describe the deformation of the midsurface: the elongations ϵ_α and the changes of curvature κ_α in the two coordinate directions, the in-plane shear deformation ω, and the torsional (twist) deformation τ. The analogous expressions for a fluid shell are obtained by applying the transformation eqn [150] to eqns [216] and [217].

The equations given above include all the special cases commonly considered in geodynamics. The first is that of a flat ($R_1=R_2=\infty$, $A_1=A_2=1$) elastic plate with constant thickness h. The (uncoupled) equations governing the flexural and membrane modes are (Landau and Lifshitz, 1986)

$$\frac{Eh^3}{12(1-\nu^2)}\nabla_\mathrm{h}^4\zeta_3=P_3 \quad [218a]$$

$$\frac{Eh}{2(1-\nu^2)}[(1-\nu)\nabla_\mathrm{h}^2\boldsymbol{\zeta}+(1+\nu)\nabla_\mathrm{h}(\nabla_\mathrm{h}\cdot\boldsymbol{\zeta})] =-P_1\mathbf{e}_1-P_2\mathbf{e}_2 \quad [218b]$$

where $\boldsymbol{\zeta}$ is the 2-D (in-plane) displacement vector, ζ_3 is the normal displacement, and $\mathbf{P}$ is the 3-D load vector. The quantity $Eh^3/12(1-\nu^2)\equiv D$ is called the flexural rigidity. Equation [218a] has been widely used to model the deformation of the lithosphere caused by topographic loading (e.g., Watts, 1978). McKenzie and Bowin (1976) generalized eqn [218a] to a thick incompressible plate that need not be thin relative to the load wavelength.

A second important special case is that of a spherical shell with radius R. If $\theta_1=\theta$ (colatitude) and $\theta_2=\phi$ (longitude), then $A_1=R$ and $A_2=R\sin\theta$. Turcotte (1974) used these equations in the membrane limit to estimate the magnitude of the stresses generated in the lithosphere when it moves relative to the Earth's equatorial bulge and found that they may be large enough to cause propagating fractures. Tanimoto (1998) used the complete (membrane plus flexural) equations to estimate the displacement and state of stress in subducting lithosphere and concluded that the state of stress is strongly influenced by the spherical geometry. Mahadevan et al. (2010) used scaling analysis and numerical solutions of the complete equations to investigate the causes of the curvature and segmentation of subduction zones.

A third case is pure membrane flow in a flat viscous sheet. A well-known application is the thin-sheet model for continental deformation of England and McKenzie (1983), whose eqn [16] is just the viscous analogue of eqn [218b] with a power-law rheology of the form [43f] and expressions for P_1 and P_2 representing the lateral forces arising from variations in crustal thickness. Schmalholz (2011) used a simplified one-dimensional version of the same equation to study the gravity-driven necking instability (slab breakoff) of a vertical sheet with a power-law rheology.

A fourth limiting case is the finite-amplitude deformation of a 2-D fluid sheet. By symmetry, $\partial_2=u_2=K_2=\epsilon_2=\omega=\kappa_2=\tau=S=H=0$ for this case. If $\theta_1\equiv s$ is the arclength along the midsurface, then $A_1=A_2=1$. The instantaneous midsurface velocity $\mathbf{u}(s,t)$ produced by a given loading distribution $P_i(s,t)$ is governed by the fluid analogues of eqns [214]–[217], while the evolution of the midsurface position $\mathbf{x}_0(s,t)$ and the sheet thickness $h(s,t)$ is described by the kinematical equations

$$D_t\mathbf{x}_0=\mathbf{u},\quad D_th=-h\epsilon_1 \quad [219]$$

where D_t is a convective derivative that follows the motion of material points on the (stretching) midsurface (Buckmaster et al., 1975). Ribe (2003) used these equations to determine scaling laws for the periodic buckling instability of a viscous sheet falling onto a horizontal surface, which are consistent with the apparent widening of subducted lithosphere imaged beneath some subduction zones (Ribe et al., 2007) and with buckling instabilities in 2-D numerical simulations of

subduction in a mantle with a viscosity increase at 660 km depth (Lee and King, 2011). Buffett (2006) used the same equations to derive an expression for the bending-induced horizontal force that resists the slab pull on a subducting plate. Buffett and Becker (2012) used an extended form of the equations with the power-law rheology [43f] to conclude that the effective power-law exponent of subducting lithosphere is $n \approx 14$.

A fifth limiting case of interest is the deformation of a viscous sheet immersed in a second fluid with a lower density and viscosity. The buoyancy force that drives the motion is now balanced by a combination of the viscous tractions exerted by the external fluid (which are formally included in the load terms P_i in eqn [214]) and the internal viscous force that resists the deformation of the sheet by stretching and bending. A scaling analysis (Ribe, 2010) shows that the fundamental dimensionless quantity that controls the (bending-dominated) dynamics of a sheet of thickness h and viscosity η_2 immersed in a fluid with viscosity $\eta_1 \ll \eta_2$ is the stiffness

$$\mathrm{St} = \frac{\eta_2}{\eta_1}\left(\frac{h}{\ell_b}\right)^3 \qquad [220]$$

where the bending length ℓ_b is the length of the portion of the sheet where internal bending stresses are significant. The driving buoyancy force is balanced primarily by the external tractions if $\mathrm{St} < 1$ (Stokes sinking regime), and by the internal bending resistance if $\mathrm{St} \gg 1$ (flexural regime). Ribe (2010) and Li and Ribe (2012) verified the pertinence of the parameter St by means of numerical (BEM) solutions for two- and three-dimensional viscous sheets subducting freely in a less viscous ambient mantle.

Finally, Medvedev and Podladchikov (1999) have derived a higher-order extended thin-sheet approximation (ETSA) that includes layer-parallel shear (lubrication flow) in addition to deformation by stretching and bending. The ETSA allows the viscosity η to vary both along and across the sheet, but is limited to nearly flat sheets.

7.04.8.5 Effective Boundary Conditions from Thin-Layer Flows

The interaction of a convective flow with a passive and rheologically distinct lithosphere can be studied using a simple extension of thin-sheet theory in which the sheet's dynamics are reduced to an equivalent boundary condition on the underlying flow. For simplicity, consider a flat fluid sheet with constant thickness h and laterally variable viscosity $\hat{\eta}(x_1, x_2)$, overlying a mantle whose viscosity just below the sheet is $\overline{\eta}(x_1, x_2) \ll \hat{\eta}$. In the following, superposed hats and bars denote quantities within the sheet and in the mantle just below it, respectively.

Because $\hat{\eta} \gg \overline{\eta}$, it is consistent to assume that the flow within the sheet is the sum of a dominant membrane component plus a small shear component that is required to balance the shear stress applied by the convective flow beneath. The dominance of membrane flow means that the lateral velocities $\hat{u}_\alpha$ and the pressure $\hat{p}$ are independent of the depth x_3 to lowest order, where $x_3 = 0$ and $x_3 = -h$ are the upper and lower surfaces of the sheet, respectively. Because the sheet is passive (nonbuoyant), the lateral force balance within it is $\partial_\alpha \hat{\sigma}_{\alpha\beta} + \partial_3 \hat{\sigma}_{\beta 3} = 0$. Integrating this across the sheet subject to the free-surface condition $\hat{\sigma}_{\beta 3}|_{z=0} = 0$, we obtain

$$h\partial_\alpha\left[-\hat{p}\delta_{\alpha\beta} + 2\hat{\eta}\hat{e}_{\alpha\beta}\right] - \hat{\sigma}_{\beta 3}|_{z=-h} = 0 \qquad [221]$$

where $e_{\alpha\beta} = (\partial_\alpha u_\beta + \partial_\beta u_\alpha)/2$ is the strain rate tensor. Now, continuity of the velocity and stress at $x_3 = -h$ requires $\hat{u}_i = \overline{u}_i$, $\hat{e}_{\alpha\beta} = \overline{e}_{\alpha\beta}$, $\hat{\sigma}_{\beta 3}|_{z=-h} = \overline{\sigma}_{\beta 3}$, and $-\hat{p} + 2\hat{\eta}\hat{e}_{33} \equiv -\hat{p} - 2\hat{\eta}\overline{e}_{\lambda\lambda} = \overline{\sigma}_{33}$, where the continuity equation $\hat{e}_{33} = -\hat{e}_{\lambda\lambda}$ has been used. Substituting these relations into eqn [221], we obtain an effective boundary condition that involves only mantle (barred) variables and the known viscosity $\hat{\eta}$ of the sheet:

$$h^{-1}\overline{\sigma}_{\beta 3} = 2\partial_\alpha\left[\hat{\eta}\left(\overline{e}_{\alpha\beta} + \overline{e}_{\lambda\lambda}\delta_{\alpha\beta}\right)\right] + \partial_\beta \overline{\sigma}_{33} \qquad [222]$$

Equation [222] remains valid even if the sheet has a power-law rheology [43f], because $\hat{I} \approx \left(\hat{e}_{\alpha\beta}\hat{e}_{\alpha\beta} + \hat{e}_{\lambda\lambda}^2\right)^{1/2}$ can be written in terms of the mantle variables using the matching condition $\hat{e}_{\alpha\beta} = \overline{e}_{\alpha\beta}$.

In reality, the sheet thickness is governed by the conservation law $\partial_t h = -\partial_\alpha(h\hat{u}_\alpha)$ and will therefore not in general remain constant. However, this effect is ignored in most applications. Weinstein and Olson (1992) used the 1-D version of eqn [222] with a power-law rheology for $\hat{\eta}$ to study the conditions under which a highly non-Newtonian sheet above a vigorously convecting Newtonian fluid exhibits platelike behavior. Ribe (1992) used the spherical-coordinate analogue of eqn [222] to investigate the generation of a toroidal component of mantle flow by lateral viscosity variations in the lithosphere.

Although the foregoing discussion was restricted to passive high-viscosity ($\hat{\eta} \gg \overline{\eta}$) sheets, flows in low-viscosity layers can also be reduced to effective boundary conditions by exploiting the fact that they are described by lubrication theory. An example is Morris (2008), who derived an effective boundary condition for the influence of a thin low-viscosity layer on the large-scale flow in a channel driven by a shear stress applied to its end.

7.04.8.6 Solitary Waves

Another flow that can be studied using a long-wave approximation is the motion of finite-amplitude solitary waves in a cylindrical conduit of low-viscosity fluid embedded in a fluid of higher viscosity (Olson and Christensen, 1986; Scott et al., 1986). These waves are examples of kinematic waves (Whitham, 1974), which occur when there is a functional relation between the density or amplitude of the medium and the flux of a conserved quantity like mass; other examples include the propagation of a pulse of wastewater down a gutter and the flow of traffic on a crowded highway.

To illustrate the basic ideas, we follow here the derivation of Olson and Christensen (1986). Consider an infinite vertical conduit with a circular cross section of radius $R(z, t)$ and area $A(z, t)$, where z is the height and t is time. The conduit contains fluid with density ρ_L and viscosity η_L and is embedded in an infinite fluid with density $\rho_M \equiv \rho_L + \Delta\rho$ and viscosity $\eta_M \gg \eta_L$. Conservation of mass in the conduit requires

$$A_t + Q_z = 0 \qquad [223]$$

where $Q(z, t)$ is the volume flux and subscripts indicate partial derivatives. Because $\eta_M \gg \eta_L$, the fluid in the conduit sees

the wall as no-slip. The volume flux is therefore given by Poiseuille's law

$$Q = -\frac{A^2}{8\pi\eta_L} P_z \qquad [224]$$

where the nonhydrostatic pressure P in the conduit is determined by the requirement that the normal stress σ_{rr} in the radial (r) direction be continuous at $r = R$. Now, the pressure in the matrix is hydrostatic, and the deviatoric component of σ_{rr} in the conduit is negligible relative to that in the matrix because $\eta_M \gg \eta_L$. Continuity of σ_{rr} therefore requires $\rho_L g z - P = \rho_M g z + 2\eta_M u_r(R)$. However, because the flow in the matrix is dominantly radial, the continuity equation requires $u \propto r^{-1}$, whence $u_r(R) = -u(R)/R = -A_t/2A$ and

$$P = -\Delta\rho g z + \eta_M A^{-1} A_t \qquad [225]$$

Substituting eqn [225] into eqn [224] and using eqn [223], we obtain

$$Q = \frac{A^2}{8\pi\eta_L}\left[\Delta\rho g + \eta_M \left(A^{-1} Q_z\right)_z\right] \qquad [226]$$

The nonlinear coupled eqns [223] and [226] admit finite-amplitude traveling wave solutions wherein $A = A(z - ct)$ and $Q = Q(z - ct)$, where c is the wave speed. Substitution of these forms into eqns [223] and [226] yields a dispersion relation for c as a function of the minimum and maximum amplitudes of the wave (Olson and Christensen, 1986, eqn [22]). The most interesting case is that of an isolated solitary wave that propagates without change of shape along an otherwise uniform conduit with $A = A_0$ and $Q = \Delta\rho g A_0^2/8\pi\eta_L \equiv Q_0$. In the limit of large amplitude, the velocity of such a wave is

$$c = 2c_0 \ln\left(A_{\max}/A_0\right) \qquad [227]$$

where $c_0 \equiv Q_0/A_0$ is the average Poiseuille velocity in the conduit far from the wave and $A_{\max}$ is the maximum cross-sectional area of the wave. The speed of the wave thus increases with its amplitude. Geophysical applications of conduit solitary waves include Whitehead and Helfrich's (1988) suggestion that such waves might transport deep mantle material rapidly to the surface with little diffusion or contamination. Schubert et al. (1989) subsequently showed numerically that solitary waves can also propagate along the conduits of thermal plumes in a fluid with temperature-dependent viscosity. Such solitary waves have been invoked to explain the origin of the 'V-shaped' topographic ridges on the ocean floor south of Iceland (Albers and Christensen, 2001; Ito, 2001).

Historically, the potential importance of solitary waves in geodynamics was first demonstrated theoretically using equations describing the migration of a low-viscosity melt phase in a deformable porous matrix (McKenzie, 1984; see also **Chapter 7.02**.) Even though the mathematical description of melt/matrix solitary waves does not involve a long-wave approximation, it is appropriate to discuss them here on account of their strong physical analogies with conduit solitary waves. The existence of solitary-wave solutions to the equations governing two-phase porous flow was first demonstrated by Richter and McKenzie (1984) and Scott and Stevenson (1984). The mathematical properties of such waves were further investigated by Barcilon and Richter (1986), who concluded that they are probably not solitons, that is, that they do not possess an infinite number of conservation laws. Scott and Stevenson (1986) and Barcilon and Lovera (1989) showed analytically that 1-D solitary waves are unstable to 2-D perturbations. Wiggins and Spiegelman (1993) subsequently showed numerically that both 1-D and 2-D waves are unstable to 3-D perturbations, which suggests that melt/matrix solitary waves should be quasispherical 3-D structures. An important analogy between melt/matrix and conduit solitary waves was demonstrated by Whitehead and Helfrich (1986), who showed that the governing equations in both cases reduce to the Korteweg–de Vries equation in the limit of small amplitude.

7.04.9 Hydrodynamic Stability and Thermal Convection

Not every correct solution of the governing equations of fluid mechanics can exist in nature or in the laboratory. To be observable, the solution must also be stable, that is, able to maintain itself against the small disturbances or perturbations that are ubiquitous in any physical environment. Whether this is the case can be determined by linear stability analysis, wherein one solves the linearized equations that govern infinitesimal perturbations of a solution of interest (the basic state) to determine the conditions under which these perturbations grow (instability) or decay (stability). Typically, one expands the perturbations in normal modes that satisfy the equations and boundary conditions and then determines which modes have a rate of exponential growth with a positive real part. However, linear stability analysis describes only the initial growth of perturbations and is no longer valid when their amplitude is sufficiently large that nonlinear interactions between the modes become important. Various nonlinear stability methods have been developed to describe the dynamics of this stage.

Here, we illustrate these methods for two geodynamically important instabilities: the R–T instability of a buoyant layer and Rayleigh–Bénard (R–B) convection between isothermal surfaces. The section concludes with a discussion of how similar methods have been applied to thermal convection in more complex and realistic systems.

7.04.9.1 Rayleigh–Taylor Instability

We have already encountered the R–T instability in **Section 7.04.4.4**, where it served as an example of how intermediate asymptotic limits of a function can be identified and interpreted. We now outline the linear stability analysis that leads to expression [38] for the growth rate of infinitesimal perturbations.

The model geometry is shown in **Figure 2(a)**. The flow in both fluids satisfies the Stokes equations $\nabla p = \eta\nabla^2\mathbf{u}$, which must be solved subject to the appropriate boundary and matching conditions. Let the velocities in the two layers be $\mathbf{u}_n \equiv (u_n, v_n, w_n)$. The vanishing of the normal velocity and the shear stress at $z = -h$ requires

$$w_1(-h) = \partial_z u_1(-h) = \partial_z v_1(-h) = 0 \quad [227]$$

where the arguments x_1, x_2, and t of the variables have been suppressed. The velocity must vanish at $z = \infty$, which requires

$$u_2(\infty) = v_2(\infty) = w_2(\infty) = 0 \quad [228]$$

Finally, continuity of the velocity and traction at the interface requires

$$[\mathbf{u}\cdot\mathbf{t}] = [\mathbf{u}\cdot\mathbf{n}] = [\mathbf{t}\cdot\boldsymbol{\sigma}\cdot\mathbf{n}] = [\mathbf{n}\cdot\boldsymbol{\sigma}\cdot\mathbf{n}] + g\Delta\rho\zeta = 0 \quad [230]$$

where $[\ldots]$ denotes the jump in the enclosed quantity from fluid 1 to fluid 2 across the interface $z = \zeta$, $\boldsymbol{\sigma}$ is the nonhydrostatic stress tensor,

$$\mathbf{n} = \left(1 + |\nabla_h\zeta|^2\right)^{-1/2}(\mathbf{e}_z - \nabla_h\zeta) \quad [231]$$

is the unit vector normal to the interface, and $\mathbf{t}$ is any unit vector tangent to the interface. Because the hydrostatic pressure gradients $-\mathbf{e}_z g\rho_n$ in the two fluids are different, the nonhydrostatic part of the normal stress jumps by an amount $-g\Delta\rho\zeta$ across the interface even though the total normal stress is continuous there. The final relation required is the kinematic condition

$$\partial_t\zeta + \mathbf{u}(\zeta)\cdot\nabla_h\zeta = w(\zeta) \quad [232]$$

which expresses the fact that the interface $z = \zeta$ is a material surface.

Because the Stokes equations are linear, the nonlinearity of the problem resides entirely in the conditions [230] and [232]. To linearize them, we expand all field variables in Taylor series about the undisturbed position $z = 0$ of the interface and eliminate all terms of order ζ^2 and higher, noting that $\mathbf{u} \propto \zeta$ because the flow is driven entirely by the buoyancy associated with the disturbance of the interface. The resulting linearized matching and kinematic conditions are

$$[u] = [v] = [w] = [\sigma_{xz}] = [\sigma_{yz}] = [\sigma_{zz}] + g\Delta\rho\zeta = 0 \quad [233]$$

$$\partial_t\zeta = w(0) \quad [234]$$

To lowest order in ζ, the flow is purely poloidal (Ribe, 1998) and can be described by two poloidal potentials $\mathcal{P}_n$ (one for each of the layers $n = 0$ and 1) that satisfy $\nabla_h^2\nabla^4\mathcal{P}_n = 0$ (see Section 7.04.5.2.2). Let the (exponentially growing) deformation of the interface be $\zeta = \zeta_0 f(x,y)e^{st}$, where s is the growth rate and f is a planform function satisfying $\nabla_h^2 f = -k^2 f$, where $k \equiv |\mathbf{k}|$ is the magnitude of the horizontal wavevector $\mathbf{k}$. Then, the solutions for $\mathcal{P}_n$ must have the form

$$\mathcal{P}_n = P_n(z)f(x,y)e^{st} \quad [235]$$

Substitution of eqn [235] into eqns [55], [228], [229], and [233] yields the two-point BVP

$$\left(D^2 - k^2\right)^2 P_n = 0 \quad [236a]$$

$$P_1(-h) = D^2P_1(-h) = P_2(\infty) = DP_2(\infty) = 0 \quad [236b]$$

$$[P] = [DP] = \left[\eta\left(D^2 + k^2\right)P\right] = \left[\eta D\left(3k^2 - D^2\right)P\right] - g\Delta\rho\zeta_0 = 0 \quad [236c]$$

where $D = \mathrm{d}/\mathrm{d}z$. The general solution of eqn [236a] is

$$P_n = (A_n + B_n z)e^{-kz} + (C_n + D_n z)e^{kz} \quad [237]$$

Substitution of eqn [237] into eqns [236b] and [236c] yields eight algebraic equations that can be solved for the constants $A_n - D_n$. The growth rate is then determined from the transformed kinematic condition $\zeta_0 s = k^2 P_1(0)$, yielding eqn [38].

The R–T instability has been used to model a variety of mantle processes. The instability of a thin layer beneath an infinite fluid half-space was studied by Selig (1965) and Whitehead and Luther (1975) and used by the latter as a model for the initiation of mantle plumes. Lister and Kerr (1989a) studied analytically the instability of a rising horizontal cylinder of buoyant fluid, motivated in part by suggestions that the R–T instability in this geometry might explain the characteristic spacing of island-arc volcanoes (Marsh and Carmichael, 1974) and of volcanic centers along mid-ocean ridges (Whitehead et al., 1984). They found that the growth rate s and the most unstable wavelength are independent of the ambient fluid/cylinder viscosity ratio γ when $\gamma \gg 1$, unlike flat layers for which $s \propto \gamma^{1/3}$. Canright and Morris (1993) performed a detailed scaling analysis of the instability for two layers of finite depth (see Section 7.04.4.4) and studied the nonlinear evolution of a Newtonian or power-law layer above an effectively inviscid half-space as a model for the initiation of subduction (see Section 7.04.8.3). Ribe (1998) used a weakly nonlinear analysis to study planform selection and the direction of superexponential growth (spouting) in the R–T instability of a two-layer system. Analytical solutions of the R–T instability with more complicated rheological and density structures have been used to model the delamination of the lowermost lithosphere (Conrad and Molnar, 1997; Houseman and Molnar, 1997; Molnar and Houseman, 2004). Lister et al. (2011) studied the R-T instability of an inclined buoyant viscous cylinder immersed in a more viscous fluid, a possible model for the formation of hotspot volcanic chains (Whitehead, 1982). Finally, Kaus and Becker (2007) used a semi-analytical method to study the effect of elasticity on the R–T instability.

7.04.9.2 Rayleigh–Bénard Convection

R–B convection in a fluid layer is the paradigmatic case of a pattern-forming instability. The classic R–B configuration (Figure 13) comprises fluid with constant kinematic viscosity ν and thermal diffusivity k confined between horizontal planes $z = \pm d/2$ held at temperatures $\mp\Delta T/2$. The fluid density depends on temperature as $\rho = \rho_0(1 - \alpha T)$, where α is the coefficient of thermal expansion. The planes $z = \pm d/2$ may be either rigid or free-slip, and k is the characteristic horizontal wavenumber of the convection pattern. In the rest of this subsection, all variables will be nondimensionalized using d, d^2/κ, κ/d, and ΔT as scales for length, time, velocity, and temperature, respectively, and $\{x_1, x_2, x_3\} \equiv \{x, y, z\}$ (Figure 13).

The basic state is a motionless ($u_i = 0$) layer with a linear (conductive) temperature profile $T = -z$. Because the viscosity is constant, buoyancy forces generate a purely poloidal flow

$$u_i = \mathcal{L}_i\mathcal{P}, \quad \mathcal{L}_i = \delta_{i3}\nabla^2 - \partial_i\partial_3 \quad [238]$$

where $\mathcal{P}$ is the poloidal potential defined by eqn [53]. In the geodynamically relevant limit of negligible inertia, the equations satisfied by $\mathcal{P}$ and the perturbation θ of the conductive temperature profile are

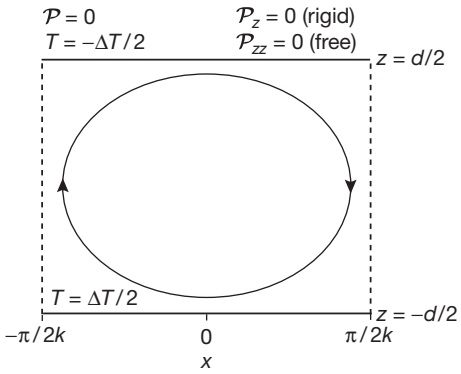


Figure 13 Geometry of Rayleigh–Bénard convection. Fluid with constant kinematic viscosity ν, thermal diffusivity κ, and thermal expansivity α is confined between horizontal planes $z = \pm d/2$ held at temperatures $\mp\Delta T/2$. The characteristic horizontal wave number of the convection pattern is k. For the special case of 2-D rolls, the roll axis is parallel to the y-direction. The boundary conditions on the poloidal scalar $\mathcal{P}$ for free and rigid boundaries are indicated above the top boundary, and subscripts denote partial differentiation.

$$\nabla^4\mathcal{P} = -\mathrm{Ra}\,\theta \qquad [239a]$$

$$\theta_t + \mathcal{L}_j\mathcal{P}\partial_j\theta - \nabla_\mathrm{h}^2\mathcal{P} = \nabla^2\theta \qquad [239b]$$

where $\nabla_\mathrm{h}^2 = \partial_{11}^2 + \partial_{22}^2$ and

$$\mathrm{Ra} = \frac{g\alpha d^3\Delta T}{\nu\kappa} \qquad [240]$$

is the Rayleigh number. The operator ∇_h^2 that appears in the primitive eqn [55] for $\mathcal{P}$ is omitted in eqn [55] because it corresponds to a simple multiplication by $-k^2$ when $\mathcal{P}$ is a periodic function of (x, y) with wavenumber k , as in the solutions to be discussed. The required boundary conditions on $z = \pm d/2$ are $\theta = 0$ together with the conditions [56] on $\mathcal{P}$.

7.04.9.2.1 Linear stability analysis

The initial-/boundary-value problem describing the evolution of small perturbations to the basic state is obtained by linearizing eqn [239] about $(\mathcal{P}, \theta) = (0, 0)$, reducing the resulting equations to a single equation for $\mathcal{P}$ by cross differentiation, and recasting the boundary conditions $\theta(\pm 1/2) = 0$ in terms of $\mathcal{P}$ with the help of eqn [239a]. For the (analytically simpler) free-slip boundary case, the result is

$$\left[\nabla^4\left(\nabla^2 - \partial_t\right) - \mathrm{Ra}\nabla_\mathrm{h}^2\right]\mathcal{P} = 0 \qquad [241a]$$

$$\mathcal{P} = \partial_{33}^2\mathcal{P} = \partial_{3333}^4\mathcal{P} = 0 \quad \text{at} \quad z = \pm 1/2 \qquad [241b]$$

Equations [241a] and [241b] admit normal mode solutions of the form

$$\mathcal{P} = P(z)f(x, y)\mathrm{e}^{\mathrm{st}} \qquad [242]$$

where s is the growth rate and $f(x, y)$ is the planform function satisfying $\nabla_\mathrm{h}^2 f = -k^2 f$. Substituting eqn [242] into eqn [241], we obtain

$$\left[\left(D^2 - k^2\right)^2\left(D^2 - k^2 - s\right) + \mathrm{Ra}k^2\right]P = 0 \qquad [243a]$$

$$P = D^2P = D^4P = 0 \quad \text{at} \quad z = \pm 1/2 \qquad [243b]$$

where $D = \mathrm{d}/\mathrm{d}z$. Equation [243] define an eigenvalue problem whose solution is

$$P = \sin n\pi\left(z + \frac{1}{2}\right), \quad s = \frac{Rak^2}{\left(n^2\pi^2 + k^2\right)^2} - n^2\pi^2 - k^2 \qquad [242]$$

where the index n defines the vertical wavelength of the mode. The growth rate s becomes positive when Ra exceeds a value $\mathrm{Ra}_0(k)$ that corresponds to marginal stability. This occurs first for the mode $n = 1$, for which

$$\mathrm{Ra}_0 = \frac{\left(\pi^2 + k^2\right)^3}{k^2} \qquad [243]$$

The most unstable wavenumber k_c and the corresponding critical Rayleigh number Ra_c are found by minimizing $\mathrm{Ra}_0(k)$, yielding

$$\mathrm{Ra}_c = \frac{27\pi^4}{4} \approx 657.5, \quad k_c = \frac{\pi}{\sqrt{2}} \approx 2.22 \qquad [246]$$

The corresponding results for convection between rigid surfaces must be obtained numerically (Chandrasekhar, 1981, pp. 36–42) and are $\mathrm{Ra}_c \approx 1707.8$, $k_c \approx 3.117$. The marginally stable Rayleigh number $\mathrm{Ra}_0(k)$ is shown for both free-slip and rigid surfaces in **Figure 14(a)**. For future reference, we note that eqn [244] can be written as

$$s = \left(\pi^2 + k^2\right)\frac{Ra - Ra_0}{Ra_0} \qquad [247]$$

7.04.9.2.2 Order-parameter equations for finite-amplitude thermal convection

Linear stability analysis predicts the initial growth rate of a normal mode with wavevector $\mathbf{k}$. Typically, however, many different modes have the same or nearly the same growth rate: examples include convection rolls with the same wavenumber but different orientations and rolls with the same orientation but slightly different wavenumbers within a narrowband around the most unstable wavenumber k_c. The question therefore arises: Among a set of modes with equal (or nearly equal) growth rates, which mode or combination of modes is actually realized for a given (supercritical) Rayleigh number? The answer depends on a combination of two factors: the nonlinear coupling between different modes due to the nonlinear terms in the governing equations and external biases imposed by the initial and/or boundary conditions. Linear stability analysis is powerless to help us here, and more complicated nonlinear theories are required.

An especially powerful method of this type is to reduce the full 3-D equations governing convection to 2-D equations for one or more order parameters that describe the degree of order or patterning in the system. Newell et al. (1993) identified four distinct classes of such equations, depending on whether the degree of supercriticality $\epsilon \equiv [(\mathrm{Ra} - \mathrm{Ra}_c)/\mathrm{Ra}_c]^{1/2}$ is small or of order unity and whether the horizontal spectrum of the allowable modes is discrete or (quasi-)continuous (**Table 4**). The spectrum is discrete for convection in an infinite layer with a periodic planform characterized by a single fundamental wave number k and also for a layer of finite extent $L = O(d)$ when

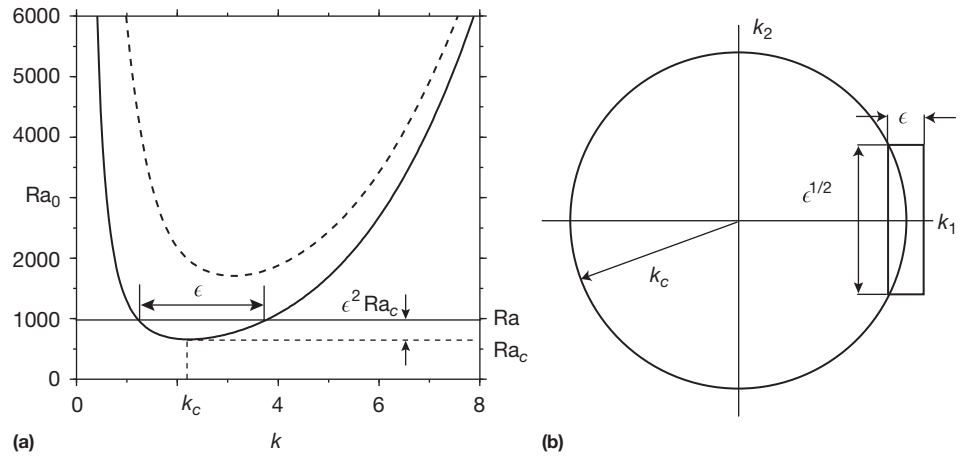


Figure 14 (a) Marginally stable Rayleigh number Ra_0 as a function of wavenumber k for Rayleigh–Bénard (R–B) convection between free-slip (solid line) and rigid (dashed line) surfaces. The critical Rayleigh number Ra_c and wavenumber k_c are indicated for the free-surface case. For a slightly supercritical Rayleigh number $\mathrm{Ra}=(1+\epsilon^2)\mathrm{Ra}_c$, a band of wavenumbers of width $\sim\epsilon$ is unstable. (b) The most unstable wavevector $\mathbf{k}=(k_1,k_2)$ for R–B convection lies on a circle of radius k_c. Changing $|\mathbf{k}|$ by an amount $\sim\epsilon$ in the vicinity of the wavevector $(k_c,0)$ for rolls with axes parallel to the x_2-direction corresponds to changing k_1 and k_2 by amounts $\sim\epsilon$ and $\sim\epsilon^{1/2}$, respectively (rectangle).

Table 4 Order-parameter equations for Rayleigh–Bénard convection

Spectrum	$\epsilon\ll1$	$\epsilon=O(1)$
Discrete	Amplitude equations	Finite-amplitude convection rolls
Continuous	Envelope equations	Slowly modulated patterns

only a few of the allowable wavenumbers $2n\pi/L$ are unstable for a given Rayleigh number. The wavenumbers of the unstable modes become ever more closely spaced as L increases, until in the limit $L/d\gg1$, they can be regarded as forming a quasi-continuous spectrum.

Historically, the first case to be studied (Malkus and Veronis, 1958) was that of a single mode with $\epsilon\ll1$. Here, the order parameter is the amplitude $A(t)$ of the dominant mode, whose temporal evolution is described by a nonlinear Landau equation (Landau, 1944). Subsequently, these weakly nonlinear results were extended to $\epsilon=O(1)$ by Busse (1967a), who used a Galerkin method to obtain numerical solutions for convection rolls and to examine their stability. The more complicated case of a continuous spectrum was first studied by Segel (1969) and Newell and Whitehead (1969), who derived the evolution equation governing the slowly varying (in time and space) amplitude envelope $A(\epsilon x,\epsilon^{1/2}y,\epsilon^2t)$ of weakly nonlinear ($\epsilon\ll1$) convection with modes contained in a narrow-band surrounding the critical wavevector $\mathbf{k}_c=(k_c,0)$ for straight parallel rolls. Finally, Newell et al. (1990) extended these results to $\epsilon=O(1)$ by deriving the phase diffusion equation that governs the slowly varying phase Φ of a convection pattern that locally has the form of straight parallel rolls.

We now examine each of these four cases in turn.

7.04.9.2.3 Amplitude equation for convection rolls

Consider convection between free-slip boundaries in the form of straight rolls with axes parallel to $\mathbf{e}_y$. The equations and boundary conditions satisfied by $\mathcal{P}$ and θ are eqn [239], where $\mathcal{P}$ and θ are independent of y and $\partial_2\equiv0$.

The basic idea of weakly nonlinear analysis is to expand the dependent variables in powers of a small parameter ϵ that measures the degree of supercriticality. By substituting these expansions into the governing equations and gathering together the terms proportional to different powers of ϵ, one reduces the original nonlinear problem to a sequence of linear (but inhomogeneous) problems that can be solved sequentially. Thus we write

$$\mathcal{P}=\sum_{n=1}^{\infty}\epsilon^n\mathcal{P}_n,\quad \theta=\sum_{n=1}^{\infty}\epsilon^n\theta_n,\quad \mathrm{Ra}=\mathrm{Ra}_0+\sum_{n=1}^{\infty}\epsilon^n\mathrm{Ra}_n \qquad [248]$$

where the last expansion can be regarded as an implicit definition of the supercriticality $\epsilon(\mathrm{Ra})$. Now, because the boundary conditions are the same on both surfaces, the hot and cold portions of the flow are mirror images of each other. The problem is therefore invariant under the transformation $\{\mathcal{P}\to-\mathcal{P},\theta\to-\theta\}$ or (equivalently) $\epsilon\to-\epsilon$. Application of the latter transformation to the expanded form of eqn [239a] shows that $\mathrm{Ra}_n=0$ for all odd n. To lowest order, therefore,

$$\mathrm{Ra}-\mathrm{Ra}_0\approx\epsilon^2\mathrm{Ra}_2 \qquad [249]$$

where $\mathrm{Ra}_2=O(1)$ is to be determined. Now, comparison of eqn [249] with eqn [247] shows that $s\propto\epsilon^2$. We therefore introduce a slow time $T\equiv\epsilon^2t$, which is of order unity during the initial stage of exponential growth, whence

$$\partial_t=\epsilon^2\partial_T \qquad [250]$$

Substituting eqns [248], [249], and [250] into eqn [239] and collecting terms proportional to ϵ, we obtain

$$\nabla^4\mathcal{P}_1+\mathrm{Ra}_0\theta_1=0,\quad \nabla^2\theta_1+\partial_{11}^2\mathcal{P}_1=0 \qquad [251]$$

subject to the boundary conditions $\mathcal{P}_1=\partial_{33}^2\mathcal{P}_1=\theta_1=0$ at $z=\pm1/2$. This is just the linear stability problem, for which the solution is

$$\mathcal{P}_1 = -\frac{C}{k^2}\cos\pi z\cos kx, \quad \theta_1 = \frac{C}{\pi^2+k^2}\cos\pi z\cos kx,$$
$$\mathrm{Ra}_0 = \frac{(\pi^2+k^2)^3}{k^2} \qquad [252]$$

where the amplitude $C(T)$ of the vertical velocity $w_1 \equiv \partial_{11}^2\mathcal{P}_1$ remains to be determined. Next, we collect terms proportional to ϵ^2 to obtain

$$\nabla^4\mathcal{P}_2 + \mathrm{Ra}_0\theta_2 = -\mathrm{Ra}_1\theta_1 \equiv 0, \qquad [253a]$$

$$\nabla^2\theta_2 + \partial_{11}^2\mathcal{P}_2 = \mathcal{L}_j\mathcal{P}_1\partial_j\theta_1 \equiv -\frac{C^2\pi}{2(\pi^2+k^2)}\sin 2\pi z. \qquad [253b]$$

Now, because the inhomogeneous term in eqn [253b] depends only on z, we must seek solutions of eqn [253] of the forms

$$\mathcal{P}_2 = \widetilde{\mathcal{P}}_2(x,z), \quad \theta_2 = \widetilde{\theta}_2(x,z) + \overline{\theta}_2(z) \qquad [254]$$

where $\widetilde{\mathcal{P}}_2$ and $\widetilde{\theta}_2$ are the fluctuating (periodic in x) parts of the solution and $\overline{\theta}_2$ is the mean (x-independent) part. The mean part of $\mathcal{P}_2$ is set to zero because the poloidal scalar is arbitrary to within an additive function of z. Now the equations satisfied by $\widetilde{\mathcal{P}}_2$ and $\widetilde{\theta}_2$ are just the homogeneous forms of eqn [253], which are identical to the order ϵ equations [251]. We may therefore set $\widetilde{\mathcal{P}}_2 = \widetilde{\theta}_2 = 0$ with no loss of generality. The solution for $\overline{\theta}_2$ is then obtained by integrating eqn [253b] subject to the boundary conditions $\overline{\theta}_2(\pm 1/2) = 0$. We thereby find

$$\mathcal{P}_2 = 0, \quad \theta_2 \equiv \overline{\theta}_2 = \frac{C^2}{8\pi(\pi^2+k^2)}\sin 2\pi z. \qquad [255]$$

Physically, $\overline{\theta}_2(z)$ describes the average heating (cooling) of the upper (lower) half of the layer that is induced by the convection. Despite appearances, eqn [255] is consistent with eqn [253a] because $\mathcal{P}_2$ is arbitrary to within an additive function of z.

The parameter Ra_2 is still not determined, so we must proceed to order ϵ^3, for which the equations are

$$\nabla^4\mathcal{P}_3 + \mathrm{Ra}_0\theta_3 = -\mathrm{Ra}_2\theta_1, \qquad [256a]$$

$$\nabla^2\theta_3 + \partial_{11}^2\mathcal{P}_3 = \partial_T\theta_1 + \mathcal{L}_j\mathcal{P}_1\partial_j\theta_2. \qquad [256b]$$

We now evaluate the RHSs of eqns [256], set $\mathcal{P}_3 = \widetilde{\mathcal{P}}_3(z,T)\cos kx$ and $\theta_3 = \widetilde{\theta}_3(z,T)\cos kx$, and reduce the resulting equations to a single equation for $\widetilde{\mathcal{P}}_3$, obtaining

$$\left[\left(D^2-k^2\right)^3 + k^2\mathrm{Ra}_0\right]\widetilde{\mathcal{P}}_3 = -\left[\frac{\mathrm{Ra}_0\left(C^3\cos 2\pi z + 4\dot{C}\right)}{4(\pi^2+k^2)} - \mathrm{Ra}_2 C\right]\cos\pi z \qquad [257]$$

where $D = \mathrm{d}/\mathrm{d}z$ and a superposed dot denotes $\mathrm{d}/\mathrm{d}T$. Now, the homogeneous form of eqn [257] is identical to the eigenvalue problem at order ϵ and thus has a solution $\propto\cos\pi z$ that satisfies the boundary conditions. The Fredholm alternative theorem then implies that the inhomogeneous eqn [257] will have a solution only if the RHS satisfies a solvability condition. In the general case, this condition is found by multiplying the inhomogeneous equation by the solution of the homogeneous adjoint problem and then integrating over the domain of the dependent variable. For our (self-adjoint) problem, the homogeneous adjoint solution is $\propto\cos\pi z$, and the procedure just described yields

$$\int_{-1/2}^{1/2}\left[\left(D^2-k^2\right)^3 + k^2\mathrm{Ra}_0\right]\widetilde{\mathcal{P}}_3\cos\pi z\,\mathrm{d}z = -\frac{\left[\mathrm{Ra}_0\left(C^3+8\dot{C}\right) - 8(\pi^2+k^2)\mathrm{Ra}_2 C\right]}{16(\pi^2+k^2)}. \qquad [258]$$

Now, the LHS of eqn [258] is identically zero, as one can easily show by integrating repeatedly by parts and applying the boundary conditions on $\widetilde{\mathcal{P}}_3$. The RHS of eqn [258] must therefore vanish. Now, by eqns [248] and [252], the amplitude of the vertical velocity is $\epsilon C \equiv W$. Because only the product of ϵ and C is physically meaningful, the definition of ϵ itself – or, equivalently, of Ra_2 – is arbitrary. Making the most convenient choice $\mathrm{Ra}_2 = \mathrm{Ra}_0$, we find that the vanishing of the RHS of eqn [258] yields the following evolution equation for the physical amplitude W:

$$\frac{\mathrm{d}W}{\mathrm{d}T} = \left(\pi^2+k^2\right)W - \frac{\mathrm{Ra}_0}{8(\mathrm{Ra}-\mathrm{Ra}_0)}W^3. \qquad [259]$$

Initially, W grows exponentially with a growth rate $W^{-1}\dot{W} = \pi^2+k^2$. At long times, however, $\dot{W}\to 0$ and the amplitude approaches a steady value

$$W(T\to\infty) = \left[\frac{8(\pi^2+k^2)(\mathrm{Ra}-\mathrm{Ra}_0)}{\mathrm{Ra}_0}\right]^{1/2}. \qquad [260]$$

7.04.9.2.4 Finite-amplitude convection rolls and their stability

The extension of the results just discussed to strongly nonlinear rolls ($\epsilon = O(1)$) between rigid surfaces is due to Busse (1967a), whose development is followed here with some changes of notation. The first step is to determine steady roll solutions of eqn [239] using a Galerkin method (Fletcher, 1984) whereby θ and $\mathcal{P}$ are expanded in orthogonal functions that satisfy the boundary conditions and the coefficients are then chosen to satisfy approximately the governing equations.

One begins by expanding θ into a complete set of Fourier modes that satisfy the boundary conditions $\theta(x,\pm 1/2) = 0$:

$$\theta = \sum_{m,n} c_{mn}\mathrm{e}^{imkx}f_n(z), \quad f_n(z) = \sin n\pi\left(z+\frac{1}{2}\right) \qquad [261]$$

where $2\pi/k$ is the wavelength of the convection rolls, $c_{mn} = \overline{c_{m-n}}$, where the overbar denotes complex conjugation, and the summations are over $-\infty < m < \infty$ and $1 \le n \le \infty$. Now, substitute eqn [261] into eqn [239a] and solve the resulting equation to obtain

$$\mathcal{P} = -\mathrm{Ra}\sum_{m,n} c_{mn}\mathrm{e}^{imkx}Q_n(mk,z) < \qquad [262a]$$

$$Q_n(r,z) = \frac{f_n(z) + n\pi h_n(r)}{\left(r^2+n^2\pi^2\right)^2}, \qquad [262b]$$

$$h_n(r) = \begin{cases} (2SC+r)^{-1}(2zC\sinh rz - S\cosh rz) & (n \text{ odd}) \\ (2SC-r)^{-1}(C\sinh rz - 2zS\cosh rz) & (n \text{ even}) \end{cases} \qquad [262c]$$

$$S = \sinh\frac{r}{2}, \quad C = \cosh\frac{r}{2}. \qquad [262d]$$

To determine the coefficients c_{mn}, we substitute eqn [262] into eqn [239b], multiply the result by $f_q(z)\mathrm{e}^{-ipkx}$, and integrate

over the fluid layer. By using for p and q all integers in the ranges of m and n, respectively, we obtain an infinite set of nonlinear algebraic equations for $c_{mn}(\mathrm{Ra})$ that can be truncated and solved numerically (Busse, 1967a).

Another advantage of the Galerkin method is that a stability analysis of the solution can easily be performed. The linearized equations governing infinitesimal perturbations $\left(\widetilde{\mathcal{P}}, \widetilde{\theta}\right)$ to the steady roll solution $(\mathcal{P}_0, \theta_0)$ are

$$\nabla^4 \widetilde{\mathcal{P}} = -\mathrm{Ra}\widetilde{\theta} \qquad [263a]$$

$$\widetilde{\theta}_t + \mathcal{L}_j \mathcal{P}_0 \partial_j \widetilde{\theta} + \partial_j \theta_0 \mathcal{L}_j \widetilde{\mathcal{P}} - \nabla_h^2 \widetilde{\mathcal{P}} = \nabla^2 \widetilde{\theta} \qquad [263b]$$

subject to $\widetilde{\theta} = \widetilde{\mathcal{P}} = \partial_3 \widetilde{\mathcal{P}} = 0$ at $z = \pm 1/2$. Now, eqns [263] are linear PDEs with coefficients that are periodic in the x-direction but independent of both z and t. The general solution can be written as a sum of solutions that depend exponentially on x, y, and t, multiplied by a function of x having the same periodicity as the stationary solution (Busse, 1967a). We therefore write

$$\widetilde{\theta} = \left\{ \sum_{m,n} \widetilde{c}_{mn} \mathrm{e}^{imkx} f_n(z) \right\} \mathrm{e}^{i(ax+by)+st} \qquad [264]$$

and note further that eqn [263a] and the boundary conditions on $\widetilde{\mathcal{P}}$ are satisfied exactly if

$$\widetilde{\mathcal{P}} = -\mathrm{Ra} \left\{ \sum_{m,n} \widetilde{c}_{mn} \mathrm{e}^{imkx} Q_n \left(\sqrt{(mk+a)^2 + b^2}, z \right) \right\} \mathrm{e}^{i(ax+by)+st} \qquad [265]$$

Substituting eqns [264] and [265] into eqn [263b], multiplying by $f_q(z)\mathrm{e}^{-i(pkx+ax+by)-st}$, and averaging over the fluid layer, we obtain an infinite system of linear equations for the coefficients $\widetilde{c}_{mn}$. The system is then truncated and the largest eigenvalue $s(\mathrm{Ra}, k, a, b)$ is determined numerically, assuming that s is real in view of the fact that $\Im(s) = 0$ at the onset of convection (Busse, 1967a).

The results show that convection rolls are stable ($s < 0$) in an elongate region of the (Ra, k) space often called the Busse balloon (Figure 15). Above $\mathrm{Ra} \approx 22600$, convection rolls of any wave number are unstable. The lower left edge of the balloon (denoted by Z in Figure 15) represents the onset of a zigzag instability in the form of rolls oblique to the original ones. Around the rest of the balloon (denoted by C), the stability of rolls is limited by the cross roll instability, which grows in the form of perpendicularly aligned rolls. The results of Busse (1967a) were extended to the case of free-slip boundaries by Straus (1972), Busse and Bolton (1984), and Bolton and Busse (1985).

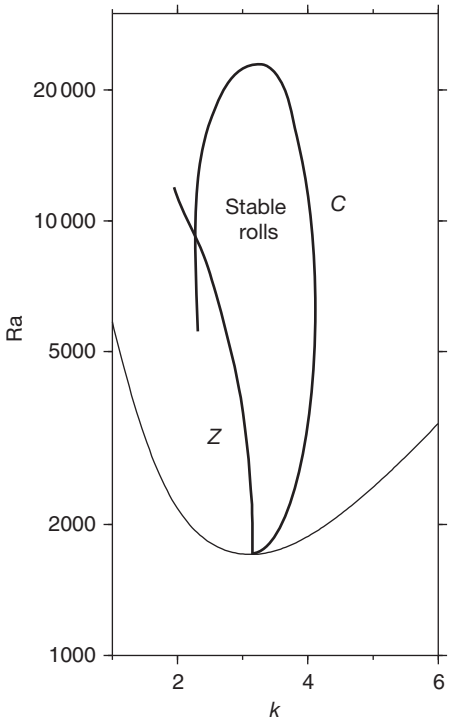


Figure 15 Region of stability in the Rayleigh number/wavenumber plane of 2-D convection rolls between rigid isothermal surfaces (the Busse balloon). Portions of the boundary of the balloon labeled Z and C correspond to the onset of the zigzag and cross roll instabilities, respectively.

7.04.9.2.5 Envelope equation for modulated convection rolls

The analyses in Sections 7.04.9.2.3 and 7.04.9.2.4 assume convection in the form of straight parallel rolls with a single dominant wavenumber k. In the laboratory, however, rolls often exhibit a more irregular pattern in which both the magnitude and direction of the wavevector $\mathbf{k} = (k_1, k_2)$ vary slowly as functions of time and position. To describe this behavior, Newell and Whitehead (1969) and Segel (1969) derived an envelope equation for modulated convection rolls whose wavevectors form a continuous spectrum within a narrowband centered on the wavevector $(k_c, 0)$ for straight parallel rolls. The envelope equation is derived via a multiscale expansion that accounts in a self-consistent way for both the fast and slow variations of the flow field in time and space. Richter (1973) applied a similar method to convection modulated by long-wavelength variations in the boundary temperatures.

The first step is to determine the scales over which the slow (i.e., long-wavelength) spatial variations occur. Because the marginal stability curve $\mathrm{Ra}_0(k)$ is a parabola in the vicinity of its minimum $(k, \mathrm{Ra}_0) = (k_c, \mathrm{Ra}_c)$, the wavenumbers $|\mathbf{k}|$ that become unstable when Ra exceeds Ra_c by an amount $\epsilon^2 \mathrm{Ra}_c$ comprise a continuous band of width $\sim\epsilon$ centered on k_c (Figure 14(a)). However, the orientation of the rolls, measured by the ratio k_2/k_1, may also vary. Now, the most unstable wavevector lies on a circle of radius k_c (Figure 14(b)). Therefore, if we change $|\mathbf{k}|$ by an amount $\sim\epsilon$ in the vicinity of the straight-roll wavevector $(k_c, 0)$, the (maximum) corresponding changes of k_1 and k_2 are $\sim\epsilon$ and $\sim\epsilon^{1/2}$, respectively. The appropriate slow variables are therefore

$$X = \epsilon x, \quad Y = \epsilon^{1/2} y, \quad T = \epsilon^2 t \qquad [266]$$

where the expression for T derives from the argument preceding eqn [250].

The essence of the multiscale procedure is to treat the flow fields (here, $\mathcal{P}$ and θ) as functions of both the fast variables (x, y, t) and the slow variables (X, Y, T). Accordingly, the asymptotic expansions analogous to eqn [248] are

$$\mathcal{P}=\sum_{n=1}^{\infty}\epsilon^n\mathcal{P}_n(X,Y,T,x,z),\quad \theta=\sum_{n=1}^{\infty}\epsilon^n\theta_n(X,Y,T,x,z). \qquad [267]$$

Because the solution for steady straight rolls is independent of y and t, $\mathcal{P}_n$ and θ_n depend on these variables only through the variables Y and T that measure the slow modulation of the pattern. By the chain rule, derivatives of the expansions [267] with respect to the fast variables transform as

$$\partial_t\to\epsilon^2\partial_T,\quad \partial_x\to\partial_x+\epsilon\partial_X,\quad \partial_y\to\epsilon^{1/2}\partial_Y,\quad \partial_z\to\partial_z. \qquad [268]$$

We now substitute eqn [267] into the governing eqns [239a] and [239b] and collect terms proportional to like powers of ϵ, just as in Section 7.04.9.2.3. The solutions at order ϵ analogous to eqn [252] are

$$\mathcal{P}_1=-\frac{1}{2k_c^2}\left[Ce^{ik_cx}+\overline{C}e^{-ik_cx}\right]\cos\pi z, \qquad [269a]$$

$$\theta_1=\frac{1}{2\left(\pi^2+k_c^2\right)}\left[Ce^{ik_cx}+\overline{C}e^{-ik_cx}\right]\cos\pi z \qquad [269b]$$

where $C=C(X,Y,T)$ is the slowly varying envelope of the roll solution and overbars denote complex conjugation. The solution at order ϵ^2 analogous to eqn [255] is

$$\mathcal{P}_2=0,\quad \theta_2=\frac{C\overline{C}}{8\pi\left(\pi^2+k_c^2\right)}\sin 2\pi z. \qquad [270]$$

The temperature θ_2 also contains free modes proportional to $\partial_X\left(C,\overline{C}\right)$ and $\partial_{YY}\left(C,\overline{C}\right)$, but these do not change the solvability condition at order ϵ^3 and can therefore be neglected. Evaluating this solvability condition as in Section 7.04.9.2.3 and setting $k_c=\pi/\sqrt{2}$, we obtain the Newell–Whitehead–Segel equation for the envelope $W\equiv\epsilon C$ of the vertical velocity field:

$$\frac{\partial W}{\partial T}-\frac{4}{3}\left(\frac{\partial}{\partial X}-\frac{i}{\sqrt{2}\pi}\frac{\partial^2}{\partial Y^2}\right)^2W=\frac{3\pi^2}{2}W-\frac{\mathrm{Ra}_c}{8(\mathrm{Ra}-\mathrm{Ra}_c)}|W|^2W. \qquad [271]$$

Equation [271] differs from eqn [259] by the addition of a diffusion-like term on the LHS, which represents the interaction of neighboring rolls via the buoyant torques they apply to each other. Equations generalizing eqn [271] to N interacting wavepackets are given by Newell and Whitehead (1969).

7.04.9.2.6 Phase diffusion equation for thermal convection

The final case is that of large-amplitude ($\epsilon=O(1)$) convection in the form of modulated rolls, which can be described by a phase diffusion equation. The derivation follows Newell et al. (1990) ≡ NPS90.

The basic assumption is that the flow field consists locally of straight parallel rolls whose wavevector varies slowly over the fluid layer. This wavevector can be written as $\mathbf{k}\equiv\nabla\phi$, where ϕ is the phase of the roll pattern. The phase is the crucial dynamical variable far from onset, because the amplitude is an algebraic function of the wavenumber and the Rayleigh number and no longer an independent parameter. The amplitude is therefore slaved to the phase gradients, and the phase diffusion equation alone suffices to describe the dynamics.

The problem involves two very different length scales, the layer depth d and the tank width $L\gg d$, and may therefore be solved using a multiscale expansion. Because the convection is strongly nonlinear, however, the relevant small parameter is no longer the degree of supercriticality (as in Section 7.04.9.2.5), but rather the inverse aspect ratio $d/L\equiv\Gamma^{-1}$. Now, the lateral scale of the variation of $\mathbf{k}$ is the tank width $L\equiv\Gamma d$, and its characteristic timescale is the lateral thermal diffusion time $L^2/\kappa\equiv\Gamma^2d^2/\kappa$. Accordingly, the slow variables that characterize the modulation of the basic roll pattern are

$$X=\Gamma^{-1}x,\quad Y=\Gamma^{-1}y,\quad T=\Gamma^{-2}t. \qquad [272]$$

The starting point of the analysis is the Galerkin representation eqns [261] and [262] for large-amplitude steady rolls with axes parallel to the y-direction and wavevector $\mathbf{k}_0=(k_0,0)$. Let A be some measure of the amplitude of this solution, which can be calculated via the Galerkin procedure as a function of Ra and k_0. Next, we seek modulated roll solutions for the dependent variables $\mathbf{v}=(\mathbf{u},\theta,p)$, where $\mathbf{u}$ is the velocity, θ is the temperature relative to the linear conductive profile, and p is the pressure. The solutions are assumed to have the form

$$\mathbf{v}=\mathbf{F}(\phi=\Gamma\Phi(X,Y,T),z;A(X,Y,T)) \qquad [273]$$

where $\mathbf{F}$ is an unknown function of the arguments indicated. In eqn [273], Φ is a slow phase variable defined such that

$$\mathbf{k}=\nabla_x\phi=\nabla_X\Phi \qquad [274]$$

where ∇_x and ∇_X are the horizontal gradient operators with respect to the variables (x,y) and (X,Y), respectively. Note that the wavevector $\mathbf{k}$ defined by eqn [274] is not required to be a small perturbation of the parallel-roll wavevector $\mathbf{k}_0$. Because derivatives act on functions of ϕ, z, X, Y, and T, the chain rule implies the transformations

$$\begin{aligned}&\partial_z\to\partial_z,\quad \nabla_x\to\mathbf{k}\partial_\phi+\Gamma^{-1}\nabla_X,\quad \partial_t\to\Gamma^{-1}(\partial_T\Phi)\partial_\phi+\Gamma^{-2}\partial_T\\&\nabla^2\to\mathrm{k}^2\partial^2_{\phi\phi}+\partial^2_{zz}+\Gamma^{-1}D\partial_\phi+\Gamma^{-2}\nabla^2_X\end{aligned} \qquad [275]$$

where $k=|\mathbf{k}|$ and $D=2\mathbf{k}\cdot\nabla_X+\nabla_X\cdot\mathbf{k}$.

Because of the slow modulation, eqn [273] is no longer an exact solution of the Boussinesq equations. We therefore seek solutions in the form of an asymptotic expansion in the powers of Γ^{-1}:

$$\mathbf{v}=\mathbf{v}_0+\Gamma^{-1}\mathbf{v}_1+\Gamma^{-2}\mathbf{v}_2+ \qquad [276]$$

where $\mathbf{v}_0$ is the Galerkin solution for steady parallel rolls. Substituting eqn [276] into the Boussinesq equations, using eqn [275], and collecting terms proportional to Γ^{-1}, we obtain a BVP for $\mathbf{v}_1$ that has a solution only if a solvability condition is satisfied. That condition yields the phase diffusion equation

$$\frac{\partial\Phi}{\partial T}+\frac{1}{\tau(k)}\nabla_X\cdot[B(k)\mathbf{k}]=0\ldots \qquad [277]$$

where the functions $\tau(k)$ and $B(k)$ are determined numerically. The diffusive character of eqn [277] becomes clear when one recalls that $\mathbf{k}=\nabla_X\Phi$. In eqn [277], an additional mean-drift advection term that is nonzero only for finite values of the Prandtl number has been neglected (cf. the first of equations (1.1) in NPS90). NPS90 show that a linear stability analysis of the straight parallel-roll solution using eqn [277]

reproduces closely the portions of the boundary of the Busse balloon (Section 7.04.9.2.4) that correspond to long-wave instabilities.

7.04.9.3 Convection at High Rayleigh Number

Even though convection in the form of steady 2-D rolls is unstable when $\mathrm{Ra} > 22600$ (Section 7.04.9.2.4), much can be learned by studying the structure of such rolls in the limit $\mathrm{Ra} \to \infty$. Because temperature variations are now confined to thin TBLs around the edges of the cells, scaling analysis and boundary-layer theory are particularly effective tools for this purpose.

7.04.9.3.1 Scaling analysis

The essential scalings for cellular convection at high Rayleigh number can be determined by a scaling analysis of the governing equations. The following derivation (for the cell geometry shown in Figure 7) generalizes the analysis of McKenzie et al (1974) to include both free-slip and rigid-surface boundary conditions. For simplicity, we assume (in this subsection only) that the aspect ratio β does not differ much from unity.

The analysis proceeds by determining six equations relating six unknown quantities: the thickness δ_p of the thermal plumes, the thickness δ_h of the horizontal TBLs, the maximum vertical velocity v_p and the vorticity ω in the plumes, the horizontal velocity u_h at the edge of the horizontal BLs, and the heat flux q across the layer (per unit length along the roll axes). The heat flux carried by an upwelling plume is

$$q \sim \rho c_p v_p \Delta T \delta_p \qquad [278]$$

where c_p is the specific heat at constant pressure. Because the convection is steady, the flux [278] must equal that lost by conduction through the top horizontal BL, implying

$$q \sim k_c d \Delta T / \delta_h \qquad [279]$$

where k_c is the thermal conductivity. The thickness δ_h of the horizontal BLs is controlled by the balance $uT_x \sim \kappa T_{yy}$ of advection and diffusion, which implies

$$u_h \Delta T / d \sim \kappa \Delta T / \delta_h^2. \qquad [280]$$

Now u_h depends on whether the horizontal surfaces are free-slip or rigid. As in our earlier scaling analysis of heat transfer from a moving sphere (Section 7.04.3.3), the velocity parallel to the boundary is constant to lowest order across a TBL at a free-slip surface (say, $n=0$) and varies linearly across a TBL at a rigid surface ($n=1$). Therefore,

$$u_h \sim v_p (\delta_h / d)^n. \qquad [281]$$

The force balance in the plumes is scaled using the vorticity equation $\nu \nabla^2 \omega = g\alpha T_x$, which implies

$$\omega \sim g\alpha \Delta T \delta_p / \nu. \qquad [282]$$

Finally, the vorticity ω in the plumes is of the same order as the rotation rate of the isothermal core, or

$$\omega \sim v_p / d. \qquad [283]$$

The simultaneous solution of eqns [278]–[283] yields scaling laws for each of the six unknown quantities as a function of R. Table 5 shows the results for v_p, δ_p, and the Nusselt number $\mathrm{Nu} \equiv d/\delta_h$. The free-slip surface heat transfer law $\mathrm{Nu} \sim \mathrm{Ra}^{1/3}$ corresponds to a dimensional heat flux $q \equiv k_c \Delta T \mathrm{Nu}/d$ that is independent of the layer depth d, whereas $q \propto d^{-2/5}$ for rigid surfaces. A revealing check of the results is to note that $v_p d/\kappa$ (column 2) is just the effective Péclet number Pe for the flow. The scaling laws for Nu (column 4) then imply $\mathrm{Nu} \sim \mathrm{Pe}^{1/(n+2)}$, in agreement with our previous expression [18] for the heat transfer from a hot sphere moving in a viscous fluid.

Table 5 Scaling laws for vigorous Rayleigh–Bénard convection

Boundaries	$v_p d/\kappa$	δ_p/d	$\mathrm{Nu} \equiv d/\delta_h$
Free-slip	$\mathrm{Ra}^{2/3}$	$\mathrm{Ra}^{-1/3}$	$\mathrm{Ra}^{1/3}$
Rigid	$\mathrm{Ra}^{3/5}$	$\mathrm{Ra}^{-2/5}$	$\mathrm{Ra}^{1/5}$

7.04.9.3.2 Flow in the isothermal core

Given the fundamental scales of Table 5, we can now carry out a more detailed analysis based on BL theory. This approach, pioneered by Turcotte (1967) and Turcotte and Oxburgh (1967) and extended by Roberts (1979), Olson and Corcos (1980), and Jimenez and Zufiria (1987) (henceforth JZ87), is applicable to convection in a layer bounded by either free-slip or rigid surfaces. Here, we consider only the former case, following JZ87 with some changes of notation.

The dimensional equations governing the flow are

$$\nu \nabla^4 \psi = -g\alpha T_x, \qquad [284a]$$

$$u\theta_x + v\theta_y = \kappa \nabla^2 T \qquad [284b]$$

where T is the temperature, $\mathbf{u} = u\mathbf{e}_x + v\mathbf{e}_y \equiv -\psi_y \mathbf{e}_x + \psi_x \mathbf{e}_y$ is the velocity, and ψ is the stream function. In the limit $\mathrm{Ra} \to \infty$, each cell comprises a nearly isothermal core surrounded by thin vertical plumes and horizontal BLs. Because temperature gradients are negligible in the core, the principal agency driving the flow within it is the shear stresses applied to its vertical boundaries by the plumes. Our first task is therefore to derive a boundary condition on the core flow that represents this agency.

Consider for definiteness the upwelling plume at the left boundary $x = -\beta d/2 \equiv x_-$ of the cell (Figure 7), and let its halfwidth be δ_p. The (upward) buoyancy force acting on the right half of this plume must be balanced by a downward shear stress at its edge $x = x_- + \delta_p$ (the shear stress on $x = x_-$ is zero by symmetry). This requires

$$\nu v_x|_{x=x_-+\delta_p} = -g\alpha \int_{x_-}^{x_-+\delta_p} (T - T_c)\,\mathrm{d}x \qquad [285]$$

where T_c ($\equiv 0$) is the temperature in the core. Multiplying eqn [285] by v_p, we obtain

$$v_p v_x|_{x=x_-+\delta_p} = -\frac{\alpha g}{\nu \rho c_p} \left[\rho c_p \int_{x_-}^{x_-+\delta_p} v_p (T - T_c)\,\mathrm{d}x \right] \qquad [286]$$

where we have used the fact that $v_p \equiv \psi_x$ is constant to first order across the plumes to take it inside the integral on the RHS. Now, the quantity [...] in eqn [286] is just the heat flux carried by the right half of the plume, which is $k_c \Delta T \mathrm{Nu}/2$. Moreover, because the plume is thin, $v_x \equiv \psi_{xx}$ can be evaluated at $x = x_-$ rather than at $x = x_- + \delta_p$. Equation [286] then becomes

$$\psi_x\psi_{xx}|_{x=x_-} = -\frac{1}{2}\frac{g\kappa\alpha\Delta T}{\nu}\mathrm{Nu} \qquad [287]$$

which is a nonlinear and inhomogeneous boundary condition on the core flow. The corresponding condition at $x=+\beta d/2$ is obtained from eqn [287] by reversing the sign of the RHS.

We now invoke the results of the scaling analysis (**Table 5**) to nondimensionalize the equations and boundary conditions for the core flow. We first write the scaling law for Nu as

$$\mathrm{Nu} = C(\beta)\mathrm{Ra}^{1/3} \qquad [288]$$

where $C(\beta)$ (to be determined) measures the dependence of the heat transfer on the aspect ratio. Rewriting the equations and boundary conditions in terms of the dimensionless variables $(x',z')=(x,z)/d$ and $\psi'=\psi/\kappa\sqrt{C(\beta)}\mathrm{Ra}^{2/3}$ and then dropping the primes, we obtain

$$\nabla^4\psi = 0, \qquad [289a]$$

$$\psi(x,\pm 1/2) = \psi_{yy}(x,\pm 1/2) = 0, \qquad [289b]$$

$$\psi(\pm\beta/2,\gamma) = \psi_x(\pm\beta/2,\gamma)\psi_{xx}(\pm\beta/2,\gamma)\mp 1/2 = 0. \qquad [289c]$$

The problem [289] can be solved either numerically (JZ87) or using a superposition method (**Section 7.04.5.4.1**).

7.04.9.3.3 TBLs and heat transfer

The next step is to determine the temperature in the plumes and the horizontal TBLs using BL theory. The temperature in all these layers is governed by the transformed BL eqn [164]. Turcotte and Oxburgh (1967) solved this equation assuming self-similarity, obtaining solutions of the form [166]. However, Roberts (1979) pointed out that this is not correct, because the fluid traveling around the margins of the cell sees a periodic boundary condition, which is alternatingly isothermal (along the horizontal boundaries) and insulating (in the plumes). Now, because eqn [164] is parabolic, its solution can be written in convolution-integral form in terms of an arbitrary boundary temperature $T_b(\tau)$ and an initial temperature profile $T(\psi,0)$ at $\tau=0$, where $\tau=\int U(s)\mathrm{d}s$ is the time-like variable introduced in **Section 7.04.7.2.1** and the velocity $U(s)$ parallel to the boundary is determined (to within the unknown scale factor $C(\beta)$) from the core flow streamfunction ψ that satisfies eqn [289]. However, in the limit $\tau\to\infty$ corresponding to an infinite number of transits around the cell, the convolution integral describing the evolution of the initial profile vanishes, and the solution is (JZ87)

$$T(\psi,\tau) = -\frac{\psi}{2(\pi\kappa)^{1/2}}\int_0^\infty \frac{T_b(\tau-t)}{t^{3/2}}\exp\left(-\frac{\psi^2}{4\kappa t}\right)\mathrm{d}t \qquad [290]$$

where all variables are dimensional. Now, T_b is known only on the top and bottom surfaces. Along the (insulating) plume centerlines, it can be found by setting to zero the temperature gradient $\partial_\psi T|_{\psi=0}$ calculated from eqn [290] and solving numerically the resulting integral equation. The temperature everywhere in the plumes and horizontal BLs can then be determined from eqn [290]. The final step is to determine the unknown scale factor $C(\beta)$ by matching the heat flow advected vertically by the plumes to the conductive heat flux across the upper boundary (JZ87, eqn [23]).

The solution of JZ87 shows that the scale factor $C(\beta)$ (and hence the Nusselt number $\mathrm{Nu}=C\mathrm{Ra}^{1/3}$) is maximum for $\beta\approx 0.75$, which may therefore be the preferred cell aspect ratio (Malkus and Veronis, 1958). A particularly interesting result from a geophysical perspective is that the temperature profile in the TBL beneath the cold upper surface exhibits a hot asthenosphere, that is, a range of depths where the temperature exceeds that of the cell's interior.

7.04.9.3.4 Structure of the flow near the corners

The solution just described breaks down near the corners of the cell, where the vorticity of the core flow solution becomes infinite and the BL approximation is no longer valid. The (dimensional) characteristic radius r_c of this corner region can be determined using a scaling argument. Consider for definiteness the corner above the upwelling plume, and let (r,ϕ) be polar coordinates with origin at the corner such that $\phi=0$ on the upper surface and $\phi=\pi/2$ on the plume centerline. Close to the corner, the stream function has the self-similar form $\psi(r,\phi)=r^\lambda F(\phi)$ for some λ, where F is given by eqn [77]. Near the corner, the dimensional forms of the boundary conditions [289b] and [289c] are

$$\psi(r,0) = \psi(r,\pi/2) = \psi_{\phi\phi}(r,0) = 0, \qquad [291a]$$

$$r^{-3}\psi_\phi(r,\pi/2)\psi_{\phi\phi}(r,\pi/2) = \frac{1}{2}\frac{g\kappa\alpha\Delta T}{\nu}\mathrm{Nu}. \qquad [291b]$$

Eqn [291b] immediately implies $\lambda=3/2$, and the stream function that satisfies all the conditions [291] is

$$\psi = \frac{\kappa}{2}\left(\frac{r}{d}\right)^{3/2}C(\beta)^{1/2}\mathrm{Ra}^{2/3}\left(\sin\frac{3}{2}\phi - \sin\frac{1}{2}\phi\right). \qquad [292]$$

Note that the vorticity $\omega\sim-(\kappa/d^2)\mathrm{Ra}^{2/3}(r/d)^{-1/2}$ becomes infinite at the corner. Now, the balance between viscous forces and buoyancy in the corner region requires $\nu\nabla^4\psi\sim -g\alpha T_x$ or

$$\psi\sim\kappa\mathrm{Ra}(r_c/d)^3. \qquad [293]$$

Equating eqn [293] with the scale $\psi\sim\kappa\mathrm{Ra}^{2/3}(r_c/d)^{3/2}$ implied by eqn [292], we find (Roberts, 1979)

$$r_c\sim\mathrm{Ra}^{-2/9}d. \qquad [294]$$

A corrected solution for the flow in the corner region that removes the vorticity singularity was proposed by JZ87.

7.04.9.4 Howard's Scaling for High-Ra Convection

While the solution for cellular convection just discussed is illuminating, we know that rolls between rigid boundaries with $R>22600$ are not stable to small perturbations (**Section 7.04.9.2.4**). In reality, convection at high Rayleigh number $\mathrm{Ra}>10^6$ is a quasiperiodic process in which the TBLs grow by thermal diffusion, become unstable, and then empty rapidly into plumes, at which point the cycle begins again. The characteristic time scale of this process can be estimated via a simple scaling argument (Howard, 1964). The thickness δ of both TBLs initially increases by thermal diffusion as $\delta\approx(\pi\kappa t)^{1/2}$. Instability sets in when the growth rate of R–T instabilities in the layers (see **Section 7.04.9.1**) becomes comparable to the thickening rate $\dot\delta/\delta$ or (equivalently) when the

Rayleigh number $\mathrm{Ra}(\delta) \equiv g\alpha\Delta T\delta^3/\nu\kappa$ based on the TBL thickness attains a critical value $\mathrm{Ra}_c \approx 10^3$. This occurs after a time

$$t_c \approx \frac{1}{\pi\kappa}\left(\frac{\kappa\nu\mathrm{Ra}_c}{g\alpha\Delta T}\right)^{2/3} \qquad [295]$$

that is independent of the layer depth. The scaling law [295] has been amply confirmed by laboratory experiments (e.g., Sparrow et al., 1970).

7.04.9.5 Thermal Convection in More Realistic Systems

As the simplest and most widely studied example of thermal convection, the R–B configuration is the source of most of what has been learned about thermal convection during the past century. From a geophysical point of view, however, it lacks many crucial features that are important in the Earth's mantle: the variation of viscosity as a function of pressure, temperature, and stress; the presence of solid-state phase changes; internal production of heat by radioactive decay; density variations associated with differences in chemical composition; and a host of other factors such as rheological anisotropy and the effect of volatile content. Because analytical methods thrive on simple model problems, the addition of each new complexity reduces their room for maneuver; but many important analytical results have been obtained nevertheless. Here, we review briefly some of these, focusing on the effects of variable viscosity, phase transitions, and chemical heterogeneity.

A variety of important analytical results have been obtained for fluids with temperature-dependent viscosity. Stengel et al. (1982) performed a linear stability analysis of convection in a fluid whose viscosity $\nu(T)$ varies strongly (by up to a factor of 2×10^4) with temperature. For fluids for which $\nu(T)$ is an exponential function, the critical Rayleigh number (based on the viscosity at the mean of the boundary temperatures) first increases and then decreases as the total viscosity contrast across the layer increases, reflecting the fact that convection begins first in a low-viscosity sublayer near the hot bottom boundary. A nonlinear stability analysis of convection in a fluid with a weak linear dependence of viscosity on temperature was carried out by Palm (1960), using a method like that described in **Section 7.04.9.2.3**. Because the variation of the viscosity breaks the symmetry between upwelling and downwelling motions, the amplitude equation analogous to eqn [259] contains a quadratic term that permits transcritical bifurcation. Subsequent work (e.g., Busse, 1967b; Palm et al., 1967) showed that the stable planforms are hexagons, hexagons and 2-D rolls, or 2-D rolls alone, depending on the Rayleigh number. Busse and Frick (1985) used a Galerkin method (**Section 7.04.9.2.4**) to study convection in a fluid whose viscosity depends strongly (but linearly) on temperature, and found that a square planform becomes stable when the viscosity contrast is sufficiently large.

Convection at high Rayleigh number in fluids with strongly variable viscosity is a particularly difficult analytical challenge, but some noteworthy results have been obtained. Morris and Canright (1984) and Fowler (1985) used boundary-layer analyses similar to that described in **Section 7.04.9.3** to determine the Nusselt number in the stagnant-lid limit $\Delta T/\Delta T_r \to \infty$ for convection in a fluid whose viscosity depends on temperature as $\nu = \nu_0 \exp[-(T-T_0)/\Delta T_r]$. Both studies find the Nusselt number to be $\mathrm{Nu} \sim (\Delta T/\Delta T_r)^{-1}\mathrm{Ra}_r^{1/5}$, where $\mathrm{Ra}_r = \alpha g\Delta T_r d^3/\nu_0\kappa$ is the Rayleigh number based on the rheological temperature scale ΔT_r and the viscosity ν_0 at the hot bottom boundary. Solomatov (1995) presented a comprehensive scaling analysis for convection in fluids with temperature- and stress-dependent viscosity. He found that three different dynamical regimes occur as the total viscosity contrast across the layer increases: quasi-isoviscous convection, a transitional regime with a mobile cold upper BL, and a stagnant-lid regime in which the upper BL is motionless and convection is confined beneath it. Busse et al. (2006) generalized Turcotte and Oxburgh's (1967) BL analysis of steady cellular convection to include thin low-viscosity layers adjoining the top and bottom boundaries, whose presence increases the cell aspect ratio that maximizes the heat transport.

The influence of a phase transition on the stability of convection in a fluid with constant viscosity was analyzed by Schubert and Turcotte (1971). The novel elements here are two matching conditions at the depth of the phase change: one of these equates the energy released by the transformation to the difference in the perturbation heat flux into and out of the phase boundary, while the second requires the phase boundary to lie on the Clapeyron slope. The resulting eigenvalue problem yields the critical Rayleigh number $\mathrm{Ra}_c(S, Ra_Q)$, where

$$S = \frac{\Delta\rho/\rho}{\alpha d(\rho g/\gamma - \beta)}, \quad \mathrm{Ra}_Q = \frac{\alpha g d^3 Q}{8 c_p \nu\kappa}, \qquad [296]$$

$\Delta\rho$ is the density difference between the phases, Q is the energy per unit mass required to change the denser phase into the lighter, γ is the Clapeyron slope, and β is the magnitude of the temperature gradient in the basic state. Ra_c is a decreasing function of S (which measures the destabilizing effect of the density change) but an increasing function of Ra_Q (which measures the stabilizing influence of latent heat). These results were extended to a divariant phase change by Schubert et al. (1972).

Convection in a chemically layered mantle was first studied by Richter and Johnson (1974), who performed a linear stability analysis of a system comprising two superposed fluid layers of equal depth and viscosity but different densities. The critical Rayleigh number Ra_c depends on the value of a second Rayleigh number $\mathrm{Ra}_\rho = g\Delta\rho d^3/\nu\kappa$ proportional to the magnitude of the stabilizing density difference $\Delta\rho$ between the layers. Three distinct modes of instability are possible: convection over the entire depth of the fluid with advection of the interface, separate convection within each layer, and standing waves on the interface, corresponding to an imaginary growth rate at marginal stability. Busse (1981) extended these results to layers of different thicknesses. Rasenat et al. (1989) examined a still larger region of the parameter space and demonstrated the possibility of an oscillatory two-layer regime with no interface deformation, in which the coupling between the layers oscillates between thermal and mechanical. Le Bars and Davaille (2002) mapped out the linear stability of two-layer convection as a function of the interlayer viscosity contrast γ and the layer depth ratio r and showed that the transition from

oscillatory to stratified convection occurs at a critical value of the buoyancy ratio $B \equiv \mathrm{Ra}_\rho/\mathrm{Ra}$ that depends on γ and r.

Another convective phenomenon that is amenable to analytical approaches is the onset of small-convection beneath the cooling oceanic lithosphere. Assuming constant viscosity, Parsons and McKenzie (1978) invoked Howard's (1964) scaling argument (**Section 7.04.9.4**) to estimate an onset time of the form [295] with $R_c = 270 - 460$ (depending on boundary conditions). Davaille and Jaupart (1994) extended that analysis to a fluid with temperature-dependent viscosity $v(T)$ and emphasized the importance of the rheological temperature scale $-v(T_0)/v'(T_0)$ where T_0 is the interior temperature below the unstable boundary layer and v' is the temperature derivative of the viscosity. Korenaga and Jordan (2003) proposed an alternative scaling analysis based on the concept of available buoyancy (the depth integral of the ratio of thermal buoyancy to viscosity; Conrad and Molnar, 1999) that predicts significantly shorter onset times. Kim and Choi (2006) used an energy stability analysis in conjunction with a similarity transformation to determine the onset time of convection in a layer of fluid with temperature-dependent viscosity cooled from above. Finally, Landuyt and Ierley (2012) performed a linear stability analysis for the onset of 2-D sublithospheric convection in a fluid with temperature-dependent viscosity subject to shear.

For the time being, studies like those just discussed represent the limit of what can be learned about convection in complex systems using analytical methods alone. The impressive additional progress that has been made using experimental and numerical approaches is discussed in **Chapters 7.03** and **7.05**.

Acknowledgments

It is a pleasure to thank U. Christensen, A. Davaille, J. Lister, and S. J. S. Morris for discussions over many years of several of the topics covered in this chapter. Detailed and constructive reviews by D. Bercovici, S. J. S. Morris, and Y. Ricard helped greatly to improve the chapter.

References

Albers M and Christensen UR (2001) Channeling of plume flow beneath mid-ocean ridges. *Earth and Planetary Science Letters* 187: 207–220.

Ansari A and Morris S (1985) The effects of a strongly temperature-dependent viscosity on Stokes's drag law: Experiments and theory. *Journal of Fluid Mechanics* 159: 459–476.

Asaadi N, Ribe NM, and Sobouti F (2011) Inferring nonlinear mantle rheology from the shape of the Hawaiian swell. *Nature* 473: 501–504.

Backus GE (1958) A class of self-sustaining dissipative spherical dynamos. *Annals of Physics* 4: 372–447.

Backus GE (1967) Converting vector and tensor equations to scalar equations in spherical co-ordinates. *Geophysical Journal of the Royal Astronomical Society* 13: 71–101.

Bai Q, Mackwell SJ, and Kohlstedt DL (1991) High-temperature creep of olivine single crystals, 1. Mechanical results for buffered samples. *Journal of Geophysical Research* 96: 2441–2463.

Barcilon V and Lovera OM (1989) Solitary waves in magma dynamics. *Journal of Fluid Mechanics* 204: 121–133.

Barcilon V and Richter FM (1986) Nonlinear waves in compacting media. *Journal of Fluid Mechanics* 164: 429–448.

Barenblatt GI (1996) *Scaling, Self-Similarity, and Intermediate Asymptotics*. Cambridge, UK: Cambridge University Press.

Batchelor GK (1967) *An Introduction to Fluid Dynamics*. Cambridge, UK: Cambridge University Press.

Batchelor GK (1970) Slender-body theory for particles of arbitrary cross-section in Stokes flow. *Journal of Fluid Mechanics* 44: 419–440.

Bercovici D (1993) A simple model of plate generation from mantle flow. *Geophysical Journal International* 114: 635–650.

Bercovici D (1994) A theoretical model of cooling viscous gravity currents with temperature-dependent viscosity. *Geophysical Research Letters* 21: 1177–1180.

Bercovici D (1998) Generation of plate tectonics from lithosphere-mantle flow and void-volatile self-lubrication. *Earth and Planetary Science Letters* 154: 139–151.

Bercovici D and Kelly A (1997) The non-linear initiation of diapirs and plume heads. *Physics of the Earth and Planetary Interiors* 101: 119–130.

Bercovici D and Lin J (1996) A gravity current model of cooling mantle plume heads with temperature-dependent buoyancy and viscosity. *Journal of Geophysical Research* 101: 3291–3309.

Biot MA (1954) Theory of stress-strain relations in anisotropic viscoelasticity and relaxation phenomena. *Journal of Applied Physics* 25: 1385–1391.

Bird RB, Armstrong RC, and Hassager O (1987) *Dynamics of Polymeric Liquids, Volume 1: Fluid Mechanics*, 2nd edn. New York, NY: John Wiley & Sons.

Blake JR (1971) A note on the image system for a Stokeslet in a no-slip boundary. *Proceedings of the Cambridge Philosophical Society* 70: 303–310.

Bloor MIG and Wilson MJ (2006) An approximate analytic solution method for the biharmonic problem. *Proceedings of the Royal Society of London A* 462: 1107–1121.

Bolton EW and Busse FH (1985) Stability of convection rolls in a layer with stress-free boundaries. *Journal of Fluid Mechanics* 150: 487–498.

Buckingham E (1914) On physically similar systems; illustrations of the use of dimensional equations. *Physical Review* 4: 345–376.

Buckmaster JD, Nachman A, and Ting L (1975) The buckling and stretching of a viscida. *Journal of Fluid Mechanics* 69: 1–20.

Buffett BA (2006) Plate force due to bending at subduction zones. *Journal of Geophysical Research* 111: B09405.

Buffett BA and Becker TW (2012) Bending stress and dissipation in subducted lithosphere. *Journal of Geophysical Research* 117, B05413.

Busse FH (1967a) On the stability of two-dimensional convection in a layer heated from below. *Journal of Mathematics and Physics* 46: 140–150.

Busse FH (1967b) The stability of finite amplitude cellular convection and its relation to an extremum principle. *Journal of Fluid Mechanics* 30: 625–649.

Busse FH (1981) On the aspect ratio of two-layer mantle convection. *Physics of the Earth and Planetary Interiors* 24: 320–324.

Busse FH and Bolton EW (1984) Instabilities of convection rolls with stress-free boundaries near threshold. *Journal of Fluid Mechanics* 146: 115–125.

Busse FH and Frick H (1985) Square-pattern convection in fluids with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 150: 451–465.

Busse FH, Richards MA, and Lenardic A (2006) A simple model of high Prandtl and high Rayleigh number convection bounded by thin low-viscosity layers. *Geophysical Journal International* 164: 160–167.

Canright DR (1987) *A Finite-Amplitude Analysis of the Buoyant Instability of a Highly Viscous Film Over a Less Viscous Half-Space*. PhD Thesis, University of California, Berkeley.

Canright D and Morris S (1993) Buoyant instability of a viscous film over a passive fluid. *Journal of Fluid Mechanics* 255: 349–372.

Chandrasekhar S (1981) *Hydrodynamic and Hydromagnetic Stability*. New York, NY: Dover.

Conrad CP and Molnar P (1997) The growth of Rayleigh-Taylor-type instabilities in the lithosphere for various rheological and density structures. *Geophysical Journal International* 129: 95–112.

Conrad CP and Molnar P (1999) Convective instability of a boundary layer with temperature- and strain-rate-dependent viscosity in terms of 'available buoyancy'. *Geophysical Journal International* 139: 51–68.

Cox RG (1970) The motion of long slender bodies in a viscous fluid Part 1. General theory. *Journal of Fluid Mechanics* 44: 791–810.

Crosby A and Lister JR (2014) Creeping axisymmetric plumes with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 745: R2.

Davaille A (1999) Two-layer thermal convection in miscible viscous fluids. *Journal of Fluid Mechanics* 379: 223–253.

Davaille A and Jaupart C (1994) Onset of thermal convection in fluids with temperature-dependent viscosity: Application to the oceanic mantle. *Journal of Geophysical Research* 99: 19,853–19,866.

Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

England P and McKenzie D (1983) Correction to: A thin viscous sheet model for continental deformation. *Geophysical Journal of the Royal Astronomical Society* 73: 523–532.

Farrell WE (1972) Deformation of the earth by surface loads. *Reviews of Geophysics and Space Physics* 10: 761–797.

Feighner M and Richards MA (1995) The fluid dynamics of plume-ridge and plume-plate interactions: An experimental investigation. *Earth and Planetary Science Letters* 129: 171–182.

Fenner RT (1975) On local solutions to non-Newtonian slow viscous flows. *International Journal of Non-Linear Mechanics* 10: 207–214.

Fleitout L and Froidevaux C (1983) Tectonic stresses in the lithosphere. *Tectonics* 2: 315–324.

Fletcher CAJ (1984) *Computational Galerkin Methods*. New York, NY: Springer Verlag.

Forte AM and Peltier WR (1987) Plate tectonics and aspherical Earth structure: The importance of poloidal-toroidal coupling. *Journal of Geophysical Research* 92: 3645–3679.

Fowler AC (1985) Fast thermoviscous convection. *Studies in Applied Mathematics* 72: 189–219.

Gantmacher FR (1960) *Matrix Theory*. 2 vols, Providence, RI: AMS Chelsea Publishing.

Gomilko AM, Malyuga VS, and Meleshko VV (2003) On steady Stokes flow in a trihedral rectangular corner. *Journal of Fluid Mechanics* 476: 159–177.

Gratton J and Minotti F (1990) Self-similar viscous gravity currents: Phase-plane formalism. *Journal of Fluid Mechanics* 210: 155–182.

Griffiths RW (1986) Thermals in extremely viscous fluids, including the effects of temperature-dependent viscosity. *Journal of Fluid Mechanics* 166: 115–138.

Hager BH, Clayton RW, Richards MA, Comer RP, and Dziewonski AM (1985) Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* 313: 541–545.

Hager BH and O'Connell RJ (1981) A simple global model of plate tectonics and mantle convection. *Journal of Geophysical Research* 86: 4843–4867.

Happel J and Brenner H (1991) *Low Reynolds Number Hydrodynamics*, 2nd edn. Dordrecht: Kluwer Academic.

Hauri EH, Whitehead JA Jr., and Hart SR (1994) Fluid dynamic and geochemical aspects of entrainment in mantle plumes. *Journal of Geophysical Research* 99: 24275–24300.

Hinch EJ (1991) *Perturbation Methods*. Cambridge, UK: Cambridge University Press.

Houseman GA and Molnar P (1997) Gravitational (Rayleigh-Taylor) instability of a layer with non-linear viscosity and convective thinning of continental lithosphere. *Geophysical Journal International* 128: 125–150.

Howard LN (1964) Convection at high Rayleigh number. In: Görtler H (ed.) *Proceedings of the Eleventh International Congress of Applied Mechanics*, pp. 1109–1115. Berlin: Springer.

Huppert HE (1982) The propagation of two-dimensional and axisymmetric viscous gravity currents over a rigid horizontal surface. *Journal of Fluid Mechanics* 121: 43–58.

Ito G (2001) Reykjanes 'V'-shaped ridges originating from a pulsing and dehydrating mantle plume. *Nature* 411: 681–684.

Jackson JD (1975) *Classical Electrodynamics*. New York, NY: John Wiley & Sons.

Jeong J-T and Moffatt HK (1992) Free-surface cusps associated with flow at low Reynolds number. *Journal of Fluid Mechanics* 241: 1–22.

Jimenez J and Zufiria JA (1987) A boundary-layer analysis of Rayleigh-Bénard convection at large Rayleigh number. *Journal of Fluid Mechanics* 178: 53–71.

Johnson RE (1980) An improved slender-body theory for Stokes flow. *Journal of Fluid Mechanics* 99: 411–431.

Katopodes FV, Davis AMJ, and Stone HA (2000) Piston flow in a two-dimensional channel. *Physics of Fluids* 12: 1240–1243.

Kaus BJP and Becker TW (2007) Effects of elasticity on the Rayleigh-Taylor instability: Implications for large-scale geodynamics. *Geophysical Journal International* 168: 843–862.

Keller JB and Rubinow SI (1976) Slender-body theory for slow viscous flow. *Journal of Fluid Mechanics* 75: 705–714.

Kerr RC and Lister JR (1987) The spread of subducted lithospheric material along the mid-mantle boundary. *Earth and Planetary Science Letters* 85: 241–247.

Kevorkian J and Cole JD (1996) *Multiple Scale and Singular Perturbation Methods*. New York, NY: Springer.

Kim MC and Choi CK (2006) The onset of buoyancy-driven convection in fluid layers with temperature-dependent viscosity. *Physics of the Earth and Planetary Interiors* 155: 42–47.

Kim S and Karrila SJ (1991) *Microhydrodynamics: Principles and Selected Applications*. Boston, MA: Butterworth-Heinemann.

Koch DM and Koch DL (1995) Numerical and theoretical solutions for a drop spreading below a free fluid surface. *Journal of Fluid Mechanics* 287: 251–278.

Koch DM and Ribe NM (1989) The effect of lateral viscosity variations on surface observables. *Geophysical Research Letters* 16: 535–538.

Korenaga J and Jordan TH (2003) Physics of multiscale convection in Earth's mantle: Onset of sublithospheric convection. *Journal of Geophysical Research* 108: 2333.

Lachenbruch AH and Nathenson M (1974) Rise of a variable viscosity fluid in a steadily spreading wedge-shaped conduit with accreting walls. *Open File Rep. 74-251*, p. 27, *United States Geological Survey*, Menlo Park, CA.

Ladyzhenskaya OA (1963) *The Mathematical Theory of Viscous Incompressible Flow*. New York, NY: Gordon and Breach.

Lamb H (1932) *Hydrodynamics*. Cambridge, UK: Cambridge University Press.

Landau LD (1944) On the problem of turbulence. *Comptes Rendus de l'Academie des Sciences de l'URSS* 44: 311–314.

Landau LD and Lifshitz E (1986) *Theory of Elasticity*, 3rd edn. Oxford: Pergamon.

Landuyt W and Ierley G (2012) Linear stability analysis of the onset of sublithospheric convection. *Geophysical Journal International* 189: 19–28.

Langlois WE (1964) *Slow Viscous Flow*. New York, NY: Macmillan.

Le Bars M and Davaille A (2002) Stability of thermal convection in two superposed miscible viscous fluids. *Journal of Fluid Mechanics* 471: 339–363.

Leahy GM and Bercovici D (2010) Reactive infiltration of hydrous melt above the mantle transition zone. *Journal of Geophysical Research* 115: B08406.

Lee C and King S (2011) Dynamic buckling of subducting slabs reconciles geological and geophysical observations. *Earth and Planetary Science Letters* 312: 360–370.

Lee SH and Leal LG (1980) Motion of a sphere in the presence of a plane interface. Part 2. An exact solution in bipolar co-ordinates. *Journal of Fluid Mechanics* 98: 193–224.

Lemery C, Ricard Y, and Sommeria J (2000) A model for the emergence of thermal plumes in Rayleigh-Bénard convection at infinite Prandtl number. *Journal of Fluid Mechanics* 414: 225–250.

Li Z and Ribe NM (2012) Dynamics of free subduction from 3-D boundary-element modeling. *Journal of Geophysical Research* 117: B06408.

Ligi M, Cuffaro M, Chierici F, and Calafato A (2008) Three-dimensional passive mantle flow beneath mid-ocean ridges: An analytical approach. *Geophysical Journal International* 175: 783–805.

Lister JR and Kerr RC (1989a) The effect of geometry on the gravitational instability of a region of viscous fluid. *Journal of Fluid Mechanics* 202: 577–594.

Lister JR and Kerr RC (1989b) The propagation of two-dimensional and axisymmetric viscous gravity currents at a fluid interface. *Journal of Fluid Mechanics* 203: 215–249.

Lister JR, Kerr RC, Russell NJ, and Crosby A (2011) Rayleigh–Taylor instability of an inclined buoyant viscous cylinder. *Journal of Fluid Mechanics* 671: 313–338.

Longman IM (1962) A Green's function for determining the deformation of the earth under surface mass loads. *Journal of Geophysical Research* 67: 845–850.

Mahadevan L, Bendick R, and Liang H (2010) Why subduction zones are curved. *Tectonics* 29: TC6002.

Malkus WVR and Veronis G (1958) Finite amplitude cellular convection. *Journal of Fluid Mechanics* 4: 225–260.

Manga M (1996) Mixing of heterogeneities in the mantle: Effect of viscosity differences. *Geophysical Research Letters* 23: 403–406.

Manga M (1997) Interactions between mantle diapirs. *Geophysical Research Letters* 24: 1871–1874.

Manga M and Stone HA (1993) Buoyancy-driven interaction between two deformable viscous drops. *Journal of Fluid Mechanics* 256: 647–683.

Manga M, Stone HA, and O'Connell RJ (1993) The interaction of plume heads with compositional discontinuities in the Earth's mantle. *Journal of Geophysical Research* 98: 19,979–19,990.

Mangler W (1948) Zusammenhang zwischen ebenen und rotationssymmetrischen Grenzschichten in kompressiblen Flüssigkeiten. *Zeitschrift für Angewandte Mathematik und Mechanik* 28: 97–103.

Marsh BD (1978) On the cooling of ascending andesitic magma. *Philosophical Transactions of the Royal Society of London* 288: 611–625.

Marsh BD and Carmichael ISE (1974) Benioff zone magmatism. *Journal of Geophysical Research* 79: 1196–1206.

McKenzie DP (1969) Speculations on the consequences and causes of plate motions. *Geophysical Journal of the Royal Astronomical Society* 18: 1–32.

McKenzie DP (1977) Surface deformation, gravity anomalies and convection. *Geophysical Journal of the Royal Astronomical Society* 48: 211–238.

McKenzie DP (1984) The generation and compaction of partially molten rock. *Journal of Petrology* 25: 713–765.

McKenzie DP and Bowin C (1976) The relationship between bathymetry and gravity in the Atlantic ocean. *Journal of Geophysical Research* 81: 1903–1915.

McKenzie DP, Roberts JM, and Weiss NO (1974) Convection in the earth's mantle: Towards a numerical simulation. *Journal of Fluid Mechanics* 62: 465–538.

Medvedev SE and Podladchikov YY (1999) New extended thin-sheet approximation for geodynamic applications – I. Model formulation. *Geophysical Journal International* 136: 567–585.

Meleshko VV (1996) Steady Stokes flow in a rectangular cavity. *Proceedings of the Royal Society of London A* 452: 1999–2022.
Mitrovica JX and Forte AM (2004) A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data. *Earth and Planetary Science Letters* 225: 177–189.
Moffatt HK (1964) Viscous and resistive eddies near a sharp corner. *Journal of Fluid Mechanics* 18: 1–18.
Molnar P and Houseman GA (2004) The effects of buoyant crust on the gravitational instability of thickened mantle lithosphere at zones of intracontinental convergence. *Geophysical Journal International* 158: 1134–1150.
Morra G, Chatelain P, Tackley P, and Koumoutsakos P (2007) Large scale three-dimensional boundary element simulation of subduction. In: Shi Y, van Albada GD, Dongarra J, and Sloot PMA (eds.) *Computational Science – ICCS 2007*, pp. 1122–1129. Berlin: Springer.
Morra G, Quevedo L, and Müller RD (2012) Spherical dynamic models of top-down tectonics. *Geochemistry, Geophysics, Geosystems* 13: Q03005.
Morra G and Regenauer-Lieb K (2006) A coupled solid-fluid method for modelling subduction. *Philosophical Magazine* 86: 3307–3323.
Morra G, Regenauer-Lieb K, and Giardini D (2006) Curvature of island arcs. *Geology* 34: 877–880.
Morris S (1980) *An Asymptotic Method for Determining the Transport of Heat and Matter by Creeping Flows with Strongly Variable Viscosity; Fluid Dynamic Problems Motivated by Island Arc Volcanism*. PhD Thesis, The Johns Hopkins University.
Morris S (1982) The effects of a strongly temperature-dependent viscosity on slow flow past a hot sphere. *Journal of Fluid Mechanics* 124: 1–26.
Morris SJS (2008) Viscosity stratification and the horizontal scale of end-driven cellular flow. *Physics of Fluids* 20: 063103.
Morris S and Canright D (1984) A boundary-layer analysis of Benard convection in a fluid of strongly temperature-dependent viscosity. *Physics of the Earth and Planetary Interiors* 36: 355–373.
Muskhelishvili NI (1953) *Some Basic Problems in the Mathematical Theory of Elasticity*. Groningen: P. Noordhoff.
Nayfeh A (1973) *Perturbation Methods*. New York, NY: John Wiley & Sons.
Newell AC, Passot T, and Lega J (1993) Order parameter equations for patterns. *Annual Review of Fluid Mechanics* 25: 399–453.
Newell AC, Passot T, and Souli M (1990) The phase diffusion and mean drift equations for convection at finite Rayleigh numbers in large containers. *Journal of Fluid Mechanics* 220: 187–252.
Newell AC and Whitehead JA Jr. (1969) Finite bandwidth, finite amplitude convection. *Journal of Fluid Mechanics* 38: 279–303.
Niordson FI (1985) *Shell Theory*. Amsterdam: North-Holland.
Novozhilov VV (1959) *Theory of Thin Shells*. Groningen: P. Noordhoff.
O'Connell RJ, Gable CW, and Hager BH (1991) Toroidal-poloidal partitioning of lithospheric plate motion. In: Sabadini R, et al. (eds.) *Glacial Isostasy, Sea Level and Mantle Rheology*, pp. 535–551. Dordrecht: Kluwer Academic Publishers.
Olson P (1990) Hot spots, swells and mantle plumes. In: Ryan MP (ed.) *Magma Transport and Storage*, pp. 33–51. New York, NY: John Wiley.
Olson P and Christensen U (1986) Solitary wave propagation in a fluid conduit within a viscous matrix. *Journal of Geophysical Research* 91: 6367–6374.
Olson P and Corcos GM (1980) A boundary-layer model for mantle convection with surface plates. *Geophysical Journal of the Royal Astronomical Society* 62: 195–219.
Olson P and Singer H (1985) Creeping plumes. *Journal of Fluid Mechanics* 158: 511–531.
Palm E (1960) On the tendency towards hexagonal cells in steady convection. *Journal of Fluid Mechanics* 8: 183–192.
Palm E, Ellingsen T, and Gjevik B (1967) On the occurrence of cellular motion in Bénard convection. *Journal of Fluid Mechanics* 30: 651–661.
Parsons B and Daly S (1983) The relationship between surface topography, gravity anomalies, and the temperature structure of convection. *Journal of Geophysical Research* 88: 1129–1144.
Parsons B and McKenzie D (1978) Mantle convection and the thermal structure of the plates. *Journal of Geophysical Research* 83: 4485–4496.
Pekeris CL (1935) Thermal convection in the interior of the Earth. *Monthly Notice of the Royal Astronomical Society: Geophysical Supplement* 3: 343–367.
Peltier WR (1974) The impulse response of a Maxwell Earth. *Reviews of Geophysics and Space Physics* 12: 649–669.
Peltier WR (1995) Mantle viscosity. In: Peltier WR (ed.) *Mantle Convection: Plate Tectonics and Global Dynamics*, pp. 389–478. New York, NY: Gordon and Breach Science Publishers.
Pozrikidis C (1990) The deformation of a liquid drop moving normal to a plane wall. *Journal of Fluid Mechanics* 215: 331–363.
Pozrikidis C (1992) *Boundary Integral and Singularity Methods for Linearized Viscous Flow*. Cambridge, UK: Cambridge University Press.
Rasenat S, Busse FH, and Rehberg I (1989) A theoretical and experimental study of double-layer convection. *Journal of Fluid Mechanics* 199: 519–540.
Rayleigh L (1922) *Theory of Sound*, vol. II, p. 313. New York, NY: Dover.
Ribe NM (1992) The dynamics of thin shells with variable viscosity and the origin of toroidal flow in the mantle. *Geophysical Journal International* 110: 537–552.
Ribe NM (1998) Spouting and planform selection in the Rayleigh-Taylor instability of miscible viscous fluids. *Journal of Fluid Mechanics* 234: 315–336.
Ribe NM (2003) Periodic folding of viscous sheets. *Physical Review E* 68: 036305.
Ribe NM (2010) Bending mechanics and mode selection in free subduction: A thin-sheet analysis. *Geophysical Journal International* 180: 559–576.
Ribe NM and Christensen U (1994) Three-dimensional modeling of plume-lithosphere interaction. *Journal of Geophysical Research* 99: 669–682.
Ribe NM and Christensen U (1999) The dynamical origin of Hawaiian volcanism. *Earth and Planetary Science Letters* 171: 517–531.
Ribe NM, Christensen UR, and Theissing J (1995) The dynamics of plume-ridge interaction, 1: Ridge-centered plumes. *Earth and Planetary Science Letters* 134: 155–168.
Ribe NM and Davaille A (2013) Dynamical similarity and density (non-) proportionality in experimental tectonics. *Tectonophysics* 608: 1371–1379.
Ribe NM and Delattre WL (1998) The dynamics of plume-ridge interaction-III. The effects of ridge migration. *Geophysical Journal International* 133: 511–518.
Ribe NM, Stutzmann E, Ren Y, and van der Hilst R (2007) Buckling instabilities of subducted lithosphere beneath the transition zone. *Earth and Planetary Science Letters* 254: 173–179.
Ricard Y, Fleitout L, and Froidevaux C (1984) Geoid heights and lithospheric stresses for a dynamic earth. *Annals of Geophysics* 2: 267–286.
Richards MA and Hager BH (1984) Geoid anomalies in a dynamic Earth. *Journal of Geophysical Research* 89: 5987–6002.
Richards MA, Hager BH, and Sleep NH (1988) Dynamically supported geoid highs over hotspots: Observation and theory. *Journal of Geophysical Research* 893: 7690–7708.
Richter FM (1973) Dynamical models for sea floor spreading. *Reviews of Geophysics and Space Physics* 11: 223–287.
Richter FM and Johnson CE (1974) Stability of a chemically layered mantle. *Journal of Geophysical Research* 79: 1635–1639.
Richter FM and McKenzie DP (1984) Dynamical models for melt segregation from a deformable matrix. *Journal of Geology* 92: 729–740.
Roberts GO (1979) Fast viscous Bénard convection. *Geophysical and Astrophysical Fluid Dynamics* 12: 235–272.
Runcorn SK (1967) Flow in the mantle inferred from the low degree harmonics of the geopotential. *Geophysical Journal of the Royal Astronomical Society* 14: 375–384.
Schmalholz SM (2011) A simple analytical solution for slab detachment. *Earth and Planetary Science Letters* 304: 45–54.
Schubert G, Olson P, Anderson C, and Goldman P (1989) Solitary waves in mantle plumes. *Journal of Geophysical Research* 94: 9523–9532.
Schubert G and Turcotte DL (1971) Phase changes and mantle convection. *Journal of Geophysical Research* 76: 1424–1432.
Schubert G, Yuen DA, and Turcotte DL (1972) Role of phase transitions in a dynamic mantle. *Geophysical Journal of the Royal Astronomical Society* 42: 705–735.
Scott DR and Stevenson DJ (1984) Magma solitons. *Geophysical Research Letters* 11: 1161–1164.
Scott DR and Stevenson DJ (1986) Magma ascent by porous flow. *Journal of Geophysical Research* 91: 9283–9296.
Scott DR, Stevenson DJ, and Whitehead JA Jr. (1986) Observations of solitary waves in a deformable pipe. *Nature* 319: 759–761.
Segel L (1969) Distant side-walls cause slow amplitude modulation of cellular convection. *Journal of Fluid Mechanics* 38: 203–224.
Selig F (1965) A theoretical prediction of salt dome patterns. *Geophysics* 30: 633–643.
Shankar PN (1993) The eddy structure in Stokes flow in a cavity. *Journal of Fluid Mechanics* 250: 371–383.
Shankar PN (2005) Eigenfunction expansions on arbitrary domains. *Proceedings of the Royal Society of London A* 461: 2121–2133.
Sleep NH (1987) Lithospheric heating by mantle plumes. *Geophysical Journal of the Royal Astronomical Society* 91: 1–11.
Solomatov VS (1995) Scaling of temperature- and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.
Sparrow EM, Husar RB, and Goldstein RJ (1970) Observations and other characteristics of thermals. *Journal of Fluid Mechanics* 41: 793–800.
Stengel KC, Oliver DS, and Booker JR (1982) Onset of convection in a variable-viscosity fluid. *Journal of Fluid Mechanics* 120: 411–431.
Stevenson DJ and Turner JS (1977) Angle of subduction. *Nature* 270: 334–336.
Stimson M and Jefferey GB (1926) The motion of two spheres in a viscous fluid. *Proceedings of the Royal Society of London A* 111: 110–116.

Stokes GG (1845) On the theories of the internal friction of fluids and of the equilibrium and motion of elastic solids. *Transactions of the Cambridge Philosophical Society* 8: 287–347.
Straus JM (1972) Finite amplitude doubly diffusive convection. *Journal of Fluid Mechanics* 56: 353–374.
Tanimoto T (1998) State of stress within a bending spherical shell and its implications for subducting lithosphere. *Geophysical Journal International* 134: 199–206.
Tovish A, Schubert G, and Luyendyk BP (1978) Mantle flow pressure and the angle of subduction: Non-Newtonian corner flows. *Journal of Geophysical Research* 83: 5892–5898.
Turcotte DL (1967) A boundary-layer theory for cellular convection. *International Journal of Heat and Mass Transfer* 10: 1065–1074.
Turcotte DL (1974) Membrane tectonics. *Geophysical Journal of the Royal Astronomical Society* 36: 33–42.
Turcotte DL and Oxburgh ER (1967) Finite amplitude convection cells and continental drift. *Journal of Fluid Mechanics* 28: 29–42.
Umemura A and Busse FH (1989) Axisymmetric convection at large Rayleigh number and infinite Prandtl number. *Journal of Fluid Mechanics* 208: 459–478.
Van Dyke M (1975) *Perturbation Methods in Fluid Mechanics*. Stanford, CA: Parabolic Press.
von Mises R (1927) Bemerkungen zur Hydrodynamik. *Zeitschrift für Angewandte Mathematik und Mechanik* 7: 425–431.
Wakiya S (1975) Application of bipolar coordinates to the two-dimensional creeping motion of a liquid. II. Some problems for two circular cylinders in viscous fluid. *Journal of the Physical Society of Japan* 39: 1603–1607.
Watts AB (1978) An analysis of isostasy in the world's oceans, part 1, Hawaiian-emperor seamount chain. *Journal of Geophysical Research* 83: 5989–6004.
Weinstein SA and Olson PL (1992) Thermal convection with non-Newtonian plates. *Geophysical Journal International* 111: 515–530.
Whitehead JA (1982) Instabilities of fluid conduits in a flowing earth – are plates lubricated by the asthenosphere? *Geophysical Journal of the Royal Astronomical Society* 70: 415–433.
Whitehead JA Jr., Dick HBJ, and Schouten H (1984) A mechanism for magmatic accretion under spreading centers. *Nature* 312: 146–148.
Whitehead JA Jr. and Helfrich KR (1986) The Korteweg-de Vries equation from laboratory conduit and magma migration equations. *Geophysical Research Letters* 13: 545–546.
Whitehead JA Jr. and Helfrich KR (1988) Wave transport of deep mantle material. *Nature* 335: 59–61.
Whitehead JA Jr. and Luther DS (1975) Dynamics of laboratory diapir and plume models. *Journal of Geophysical Research* 80: 705–717.
Whitham GB (1974) *Linear and Non-Linear Waves*. Sydney: Wiley-Interscience.
Whittaker RJ and Lister JR (2006a) Steady axisymmetric creeping plumes above a planar boundary. Part I: A point source. *Journal of Fluid Mechanics* 567: 361–378.
Whittaker RJ and Lister JR (2006b) Steady axisymmetric creeping plumes above a planar boundary. Part II: A distributed source. *Journal of Fluid Mechanics* 567: 379–397.
Whittaker RJ and Lister JR (2008a) The self-similar rise of a buoyant thermal in very viscous flow. *Journal of Fluid Mechanics* 606: 295–324.
Whittaker RJ and Lister JR (2008b) Slender-body theory for steady sheared plumes in very viscous fluid. *Journal of Fluid Mechanics* 612: 21–44.
Wiggins C and Spiegelman M (1993) Magma migration and magmatic solitary waves in 3-D. *Geophysical Research Letters* 22: 1289–1292.
Yale MM and Phipps Morgan J (1998) Asthenosphere flow model of hotspot-ridge interactions: A comparison of Iceland and Kerguelen. *Earth and Planetary Science Letters* 161: 45–56.
Yuen DA and Schubert G (1976) Mantle plumes: A boundary-layer approach for Newtonian and non-Newtonian temperature-dependent rheologies. *Journal of Geophysical Research* 81: 2499–2510.
Zhong S (1996) Analytic solutions for Stokes flow with lateral variations in viscosity. *Geophysical Journal International* 124: 18–28.

7.05 Numerical Methods for Mantle Convection

SJ Zhong, University of Colorado, Boulder, CO, USA
DA Yuen, University of Minnesota, Minneapolis, MN, USA; China University of Geosciences, Wuhan, China
LN Moresi, University of Melbourne, Melbourne, VIC, Australia
MG Knepley, University of Chicago, Chicago, IL, USA

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by S J Zhong, D A Yuen and L N Moresi, Volume 7, pp. 227–252, © 2007, Elsevier B.V.

7.05.1	**Introduction**	197
7.05.2	**Governing Equations and Initial and Boundary Conditions**	198
7.05.3	**FD, FV, and Spectral Methods**	198
7.05.3.1	Finite Difference	198
7.05.3.1.1	FD implementation of the governing equations	199
7.05.3.1.2	Approximations of spatial derivatives and solution approaches	199
7.05.3.2	FV Method	201
7.05.3.3	Spectral Methods	202
7.05.4	**An FE Method**	202
7.05.4.1	Stokes' Flow: A Weak Formulation, Its FE Implementation and Solution	203
7.05.4.1.1	A weak formulation	203
7.05.4.1.2	An FE implementation	204
7.05.4.1.3	The Uzawa algorithm for the matrix equation	205
7.05.4.1.4	Multigrid solution strategies	207
7.05.4.2	Stokes' Flow: A Penalty Formulation	208
7.05.4.3	The SUPG Formulation for the Energy Equation	209
7.05.5	**Incorporating More Realistic Physics**	210
7.05.5.1	Thermochemical Convection	210
7.05.5.1.1	Governing equations	210
7.05.5.1.2	Solution approaches	211
7.05.5.2	Solid-State Phase Transition	212
7.05.5.3	Non-Newtonian Rheology	212
7.05.5.4	Viscoelasticity	213
7.05.5.5	Mantle Compressibility	214
7.05.6	**Advanced Numerical Techniques**	214
7.05.6.1	Adaptive Mesh Refinement	215
7.05.6.2	Nonlinear Solver	215
7.05.6.3	The Preconditioned Uzawa Algorithm	215
7.05.6.4	Libraries	216
7.05.6.5	Computer Architecture	216
7.05.6.6	Boundary Element Methods	217
7.05.7	**Concluding Remarks and Future Prospect**	217
References		218

7.05.1 Introduction

The governing equations for mantle convection are derived from conservation laws of mass, momentum, and energy. The nonlinear nature of mantle rheology with its strong temperature and stress dependence and nonlinear coupling between flow velocity and temperature in the energy equation require that numerical methods be used to solve these governing equations. Understanding the dynamic effects of phase transitions (e.g., olivine to spinel phase transition) and multicomponent flow also demands numerical methods. Numerical modeling of mantle convection has a rich history since the late 1960s (e.g., McKenzie et al., 1974; Torrance and Turcotte, 1971). Great progress in computer architecture along with improved numerical techniques has helped advance the field of mantle convection into its own niche in geophysical fluid dynamics (e.g., Yuen et al., 2000).

In this chapter, we will present several commonly used numerical methods in studies of mantle convection with the primary aim of reaching out to students and new researchers in the field. First, we will present the governing equations and boundary and initial conditions for a given problem in mantle convection and discuss the general efficient strategy to solve this problem numerically (see **Section 7.05.2**). We will then briefly discuss finite-difference (FD), finite-volume (FV), and spectral methods in **Section 7.05.3**. Since finite elements (FEs) have great utility and have become very popular in the user community, we will discuss FEs in greater details as the most basic numerical tool (see **Section 7.05.4**). We will also include discussions about accelerators and the usage of libraries for massively parallel

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00130-5

computing in preparation for the coming technical advances in hardware. For simplicity and clarity, we will focus our discussion on homogeneous, incompressible fluids with the Boussinesq approximation. However, we will also describe methods for more complicated and realistic mantle situations by including compressibility, non-Newtonian rheology, solid-state phase transitions, and thermochemical (i.e., multicomponent) convection (Section 7.05.5). Finally, in Section 7.05.6, we will discuss some new developments in computational sciences, such as recent software developments, especially dealing with libraries and visualization, which may impact our future studies of mantle convection modeling in a multiscale setting.

7.05.2 Governing Equations and Initial and Boundary Conditions

The simplest mathematical formulation for mantle convection assumes incompressibility and the Boussinesq approximation (e.g., McKenzie et al., 1974). Under this formulation, the nondimensional conservation equations of the mass, momentum, and energy are (see Chapter 7.02)

$$u_{i,i} = 0 \quad [1]$$

$$\sigma_{ij,j} + RaT\delta_{i3} = 0 \quad [2]$$

$$\frac{\partial T}{\partial t} + u_i T_{,i} = \left(\kappa T_{,i}\right)_{,i} + \gamma \quad [3]$$

where u_i, σ_{ij}, T, and γ are the velocity, stress tensor, temperature, and heat production rate, respectively; Ra is a Rayleigh number; and δ_{ij} is a Kronecker delta. Repeated indexes denote summation, and $u_{,i}$ represents partial derivative of variable u with respect to coordinate x_i. These equations were obtained by using the following characteristic scales: length D, time D^2/κ; and temperature ΔT, where D is often the thickness of the mantle or a fluid layer, κ is thermal diffusivity, and ΔT is the temperature difference across the fluid layer (see Chapters 7.02 and 7.04 for discussion on nondimensionalization). The stress tensor σ_{ij} can be related to strain rate $\dot{\varepsilon}_{ij}$ via the following constitutive equation:

$$\sigma_{ij} = -P\delta_{ij} + 2\eta\dot{\varepsilon}_{ij} = -P\delta_{ij} + \eta\left(u_{i,j} + u_{j,i}\right) \quad [4]$$

where P is the dynamic pressure and η is the viscosity (see Chapter 7.02).

Substituting eqn [4] into [2] reveals three primary unknown variables: pressure, velocity, and temperature. The three governing eqns [1]–[3] are sufficient to solve these three unknowns, together with adequate boundary and initial conditions. Initial conditions are only needed for temperature due to the first-order derivative with respect to time in the energy equation. Boundary conditions are in general a combination of prescribed stress and velocity for the momentum equation and of prescribed heat flux and temperature for the energy equation. The initial and boundary conditions can be expressed as

$$T(r_i, t=0) = T_{\text{init}}(r_i) \quad [5]$$

$$u_i = g_i \text{ on } \Gamma_{g_i},\ \sigma_{ij}n_j = h_i \text{ on } \Gamma_{h_i} \quad [6]$$

$$T = T_{\text{bd}} \text{ on } \Gamma_{T_{\text{bd}}},\ (T_{,i})_n = q \text{ on } \Gamma_q \quad [7]$$

where Γ_{g_i} and Γ_{h_i} are the boundaries where ith components of velocity and forces are specified to be g_i and h_i, respectively, n_j is the normal vector of boundary Γ_{h_i}, and $\Gamma_{T_{\text{bd}}}$ and Γ_q are the boundaries where temperature and heat flux are prescribed to be T_{bd} and q, respectively.

Often, free slip (i.e., zero tangential stresses and zero normal velocities) and isothermal conditions are applied to the surface and bottom boundaries in studies of mantle dynamics, although in some studies surface velocities may be given in consistence with surface plate motions (e.g., Bunge et al., 1998). When steady-state or statistically steady-state solutions are to be sought, as they often are in mantle dynamics, the choice of initial condition can be rather arbitrary and it does not significantly affect final results in a statistical sense.

Although full time-dependent dynamics of thermal convection involve all three governing equations, an important subset of mantle dynamics problems, often termed as instantaneous Stokes' flow problem, only require solutions of eqns [1] and [2]. For Stokes' flow problem, one may consider the dynamic effects of a given buoyancy field (e.g., one derived from seismic structure) or prescribed surface plate motion on gravity anomalies, deformation rate, and stress at the surface and the interior of the mantle (Hager and O'Connell, 1981; Hager and Richards, 1989; also see Chapter 7.04).

These governing equations generally require numerical solution procedures for three reasons:

(i) The advection of temperature in eqn [3], $u_iT_{,i}$, represents a nonlinear coupling between velocity and temperature.
(ii) The constitutive law or eqn [4] is often nonlinear in that stress and strain rate follow a power-law relation. That is, the viscosity η in eqn [4] can only be considered as effective viscosity that depends on stress or strain rate.
(iii) Even for Stokes' flow problem with a linear rheology, spatial variability in viscosity can make any analytic solution method difficult and impractical.

Irrespective of numerical methods used for the treatment of the individual governing equations, it is usual to solve the coupled system explicitly in time as follows:

(i) At a given time step, solve eqns [1] and [2] (i.e., instantaneous Stokes' flow problem) for flow velocity for given buoyancy or temperature.
(ii) Update the temperature to next time step from eqn [3], using the new velocity field.
(iii) Continue this process of time stepping by going back to step (i).

7.05.3 FD, FV, and Spectral Methods

In this section, we will briefly discuss how FD, FV, and spectral methods are used in studies of mantle convection. These methods have a long history in modeling mantle convection (e.g., Gable et al., 1991; Machetel et al., 1986) and remain important in the field (e.g., Harder and Hansen, 2005; Kageyama and Sato, 2004; Stemmer et al., 2006; Tackley, 2000).

7.05.3.1 Finite Difference

FD methods have a much earlier historical beginning than FE, spectral, or FV methods because they are motivated intuitively

by differential calculus and are based on local discretization of the derivative operators based on a Taylor series expansion with an assigned order of accuracy about a given point. The unknowns at each grid point depend on those of the neighboring points from the local Taylor expansion. For examples of FD implementation, the reader is urged to consult an excellent introductory book by Lynch (2005). Additional material of Fortran 95 listings of library subprograms using FD methods can be found in Griffiths and Smith (2006). Hence, FD formulations are easy to grasp and to program, and they have been the leading tool in numerical computations since the 1950s, as evidenced by the pioneering codes written for meteorology and applied weapons research (e.g., Richtmeyer and Morton, 1967). All of the initial codes in mantle convection were written with the FD formulation (Jarvis and McKenzie, 1980; McKenzie et al., 1974; Parmentier et al., 1975; Torrance and Turcotte, 1971; Turcotte et al., 1973). The initial numerical techniques in solving mantle convection in the nonlinear regime were based on second-order FD methods (McKenzie et al., 1974; Torrance and Turcotte, 1971). In the early 1980s, splines with a fourth-order accuracy were used to solve two-dimensional (2-D) problems with variable viscosity (Christensen, 1984). Malevsky (1996) developed a three-dimensional (3-D) mantle convection code based on 3-D splines, which allowed one to reach very high Rayleigh numbers, like 10^8 (Malevsky and Yuen, 1993). Kageyama and Sato (2004) had developed a 3-D FD technique, using a baseball-like topological configuration called the 'yin–yang' grid, for solving the 3-D spherical convection problem.

7.05.3.1.1 FD implementation of the governing equations

For 2-D mantle convection within the Boussinesq approximation and isoviscous flow, a commonly used formulation employs stream function Ψ and vorticity ω to eliminate the pressure and velocities, and the governing equations are written in terms of stream function Ψ, vertical vorticity ω, and temperature T (e.g., McKenzie et al., 1974) (also see eqns [1]–[3]):

$$\nabla^2 \Psi = -\omega \qquad [8]$$

$$\nabla^2 \omega = Ra \frac{\partial T}{\partial x} \qquad [9]$$

$$\frac{\partial T}{\partial t} = \nabla^2 T + \left(\frac{\partial \Psi}{\partial x}\frac{\partial T}{\partial z} - \frac{\partial \Psi}{\partial z}\frac{\partial T}{\partial x} \right) + \gamma \qquad [10]$$

where the velocity vector is defined as

$$\vec{v} = (v_x, v_z) = \left(\frac{\partial \Psi}{\partial z}, -\frac{\partial \Psi}{\partial x} \right) \qquad [11]$$

which automatically satisfies continuity eqn [1], and ω is the only nonzero component of the vorticity vector $\nabla \times \vec{v}$, x and z are the horizontal and vertical coordinates, respectively, with z vector pointing upward, and t is the time. This formulation is convenient to be used in the FD methods (e.g., McKenzie et al., 1974). We note that eqns [8] and [9] are given by a set of coupled second-order partial differential equations. Alternatively, we could have combined them to form a single fourth-order partial differential equation in terms of Ψ, called the biharmonic equation, as developed in the numerical scheme of Christensen (1984) using bicubic splines and also by Schott and Schmeling (1998) using FDs. We note that solving the biharmonic equation by FDs takes more time than solving two coupled Laplacian equations. We have restricted ourselves here to constant viscosity for pedagogical purposes. Examples of variable viscosity convection equations can be found in Christensen and Yuen (1984) and Schubert et al. (2001).

Zero tangential stress and impermeable boundary conditions can be expressed as

$$\omega|_\Gamma = \Psi|_\Gamma = 0 \qquad [12]$$

Other boundary conditions are similar to those discussed in **Section 7.05.2**.

In 2-D mantle convection problems, one can also utilize the so-called primitive variable (i.e., velocity and dynamic pressure) formulation by solving the coupled equations of mass and momentum conservations at each time step. This procedure may involve more numerical care and computational time because of more unknowns. The examples can be found in Auth and Harder (1999) and Gerya and Yuen (2003, 2007).

For 3-D mantle convection, we can likewise employ the poloidal potential Φ and a vorticity-like scalar function Ω (Busse, 1989; Chandrasekhar, 1961; Travis et al., 1990). It is to be noted that this potential Φ is not the same as the stream function Ψ in 2-D (see **Chapter 7.04**). From the general representation of an arbitrary solenoidal vector field (Busse, 1989), we can write a 3-D velocity vector as

$$\vec{v} = \nabla \times \nabla \times \left(\vec{e}_z \Phi \right) + \nabla \times \left(\vec{e}_z \Theta \right) \qquad [13]$$

where $\vec{e}_z$ is the unit vector in the vertical, z-direction, pointing upward. Θ is the toroidal potential and is present in problems with lateral variations of viscosity (Christensen and Harder, 1991; Gable et al., 1991; Zhang and Christensen, 1993). Thus, for constant viscosity, Θ is zero unless driven by a boundary condition (e.g., Gable et al., 1991), and the velocity vector $\vec{v} = (u, v, w)$ involves higher-order derivatives of Φ in this formulation:

$$u = \frac{\partial^2 \Phi}{\partial y \partial z}, \quad v = \frac{\partial^2 \Phi}{\partial x \partial z}, \quad w = -\left(\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} \right) \qquad [14]$$

The 3-D momentum equation for constant properties can be written as a system of coupled Poisson equations in 3-D:

$$\nabla^2 \Phi = \Omega \qquad [15]$$

$$\nabla^2 \Omega = RaT \qquad [16]$$

with the energy equation as

$$\frac{\partial T}{\partial t} = \nabla^2 T - \vec{v} \cdot \nabla T + \gamma \qquad [17]$$

where all differential operators are 3-D in character and Ω is a scalar function playing a role analogous to the vorticity in the 2-D formulation.

We note that FD and FV methods using the primitive variables of velocity and pressure are currently predominant in models of 3-D mantle convection with variable viscosity (Albers, 2000; Gerya, 2010; Ismail-Zadeh and Tackley, 2010; Ogawa et al., 1991; Tackley, 1994; Trompert and Hansen, 1996).

7.05.3.1.2 Approximations of spatial derivatives and solution approaches

In FD methods, spatial derivatives in the equations mentioned earlier are approximated by differences in values between grid

points. There are many algorithms for generating the formulas in the FD approximation of spatial differential operators (e.g., Kowalik and Murty, 1993; Lynch, 2005). The most common algorithm is to use a Taylor expansion. For example, a second-order differential operator (for simplicity, consider the spatial derivative in the x-direction) can be expressed as

$$\frac{d^2 f(x)}{dx^2} = [f(x_1) + f(x_3) - 2f(x_2)]/(\Delta x)^2 \qquad [18]$$

where x_k is a grid point, while the derivative is computed at a point x, which is not necessarily a grid point (**Figure 1(a)**). This approximation is second-order accurate. We will use a recursive algorithm for generating the weights in the FD approximation with high-order accuracy. We will, for simplicity, consider the spatial derivative in the x-direction. The nth-order differential operator can be approximated by an FD operator as

$$\frac{d^n f(x)}{dx^n} = \sum_{k=1}^{j} W_{n,j,k} f(x_k) \qquad [19]$$

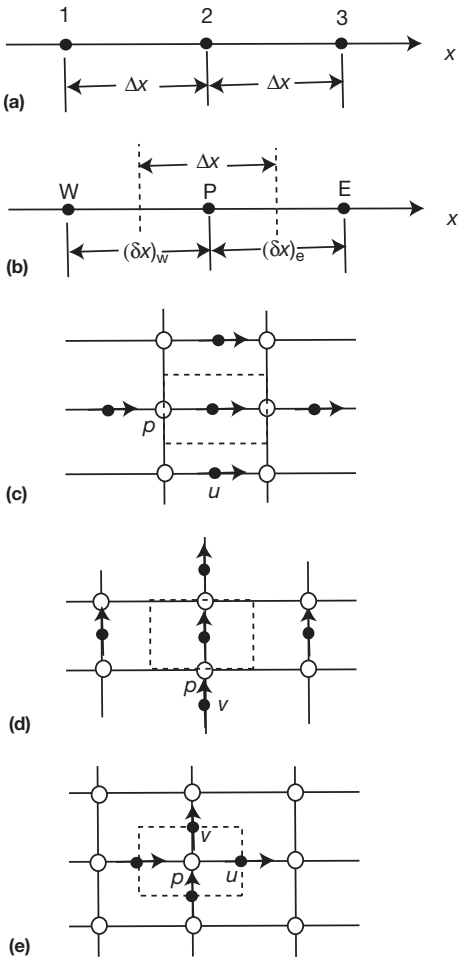


Figure 1 A simple 3-point 1-D finite-difference stencil (a), a 1-D finite-volume control cell (b), a staggered 2-D grid for velocity–pressure for horizontal (c) and vertical (d) components of the momentum equation and for the continuity equation (e).

where $W_{n,j,k}$ are the weights to be applied at grid point x_k, in order to calculate the nth-order derivative at the point x. The stencil is j grid points wide. An ideal algorithm should accurately and efficiently produce weights for any order of approximation at arbitrarily distributed grid points. Such an algorithm has been developed and tested for simple differential operators by Fornberg (1990, 1995). The cost for generating each weight is just four operations and examples of practical implementation are given by Fornberg (1995). This algorithm can be easily generalized for determining the spatial derivatives for higher-dimensional operators.

In general, a pth-order FD method converges with truncation error $O(\delta^{k-p})$ for k grid points where δ is the grid spacing. However, variable coefficients and nonlinearities, such as the heat advection term, greatly complicate the convergence. The reader can find a more detailed discussion of the issues related to high-order FD methods and spectral methods in Fornberg (1995). For mantle convection in both 2-D and 3-D configurations using this variable grid, high-order FD method has been studied by Larsen et al. (1997).

Using the stream-function/vorticity formulation for 2-D and the scalar potential approach for 3-D, we can see that we have a large system of algebraic equations to solve at each time step. The partial differential equations posed by eqns [8] and [9] for 2-D and eqns [15] and [16] for 3-D are very challenging because of their elliptical nature. The linear algebra arising from elliptical problems for large systems with variable viscosity makes them very difficult to solve accurately. They can be written down symbolically as large-scale matrix algebra problem:

$$Ku = f \qquad [20]$$

where K is a matrix operator involving the second-order spatial derivatives and the associated boundary conditions for the stream function and vorticity, u is the solution vector over the set of 2-D grid points for (Φ, Ω), and f is a vector representing the right-hand side of the system from eqns [8] and [9], namely, Ω and $\partial T/\partial x$ values at 2-D grid points from the previous time step.

The matrix equation from 2-D mantle convection problems with variable viscosity can be solved effectively and very accurately with the direct method, such as the Cholesky decomposition. The presence of strongly variable viscosity would make the matrix K poorly conditioned. The adverse condition of the matrix is further aggravated by the presence of non-Newtonian rheology, which makes the matrix K a nonlinear function of u, thus making it necessary to use iterative methods to solve the matrix equation. Multigrid iterative methods (e.g., Hackbush, 1985; Press et al., 1992; Trottenberg et al., 2001; Yavneh, 2006) are a reliable way of solving the elliptical problem with strong nonlinearities (e.g., see Davies, 1995).

For 3-D mantle convection problems, one must still employ the iterative method because of memory requirements associated with an FD grid configuration with millions of unknowns. The multigrid iterative method used for an FD grid has been used by Parmentier and Sotin (2000) for solving 3-D high Rayleigh number mantle convection. We note that for variable viscosity convection in 3-D, one must use the pressure–velocity formulation cast in FDs (e.g., Peyret and Taylor, 1983) for the momentum equation. The resulting FD equations can be solved with the multigrid method. In this connection,

higher-order FD method (Fornberg, 1995) with variable grid points together with the multigrid technique may bring about a renaissance to the FD method because of its favorable posture in terms of memory requirements over FEs.

Another source of numerical difficulties in mantle convection modeling is due to the temperature advection term, $\vec{v}\cdot\nabla T$. It is well known from linear analysis that numerical oscillations result, if simple FD schemes are used to calculate the spatial derivative due to the interaction between velocity and the FD approximations of ∇T. Excellent discussions of this problem involving numerical dispersion can be found in Kowalik and Murty (1993). Different numerical approximations have been proposed to treat the advection term in the FD; the first is by Spalding (1972), which involves a weighted upwind scheme and is first-order accurate, which has been employed in the early FD codes of mantle convection (Turcotte et al., 1973). A more popular and effective method is an iterated upwind correction scheme that is correct to second order and was proposed by Smolarkiewicz (1983, 1984). This is a scheme based on the positive definite character of the advection term. The solutions, in general, do not change much after one to two correction steps and it is easy to program for parallel computers. This scheme has been implemented by Parmentier and Sotin (2000) and by Tackley (2000) for mantle convection problems. The use of higher-order FD schemes (Fornberg, 1995) will also help to increase the accuracy of the advection of temperature because this will lead closer to a pseudospectral quality. Other advection schemes include the semi-Lagrangian technique, which is based on tracer and characteristics and has been used in mantle convection both for the stream-function method (Malevsky and Yuen, 1991) and for primitive variables (Gerya and Yuen, 2003).

7.05.3.2 FV Method

A FV method is often used to solve differential equations (Pantakar, 1980). FV methods share some common features with FE and FD methods. In FV methods, discretization of a differential equation results from integrating the equation over a control volume or control cell and approximating differential operators of reduced order at cell boundaries (Pantakar, 1980). Pantakar (1980) suggested that the FV formulation may be considered as a special case of the weighted residual method in FE in which the weighting function is uniformly one within an element and zero outside of the element. The FV method is also similar to FD method in that they both need to approximate differential operators using values at grid points.

We will use a simple example to illustrate the basic idea of the FV method (Pantakar, 1980). Consider a 1-D heat conduction equation with heat source γ:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)+\gamma=0 \qquad [21]$$

where k is the heat conductivity. This equation is integrated over a control cell bounded by dashed lines in **Figure 1(b)**, and the resulting equation is

$$\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)_e-\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)_w+\int_w^e\gamma\,\mathrm{d}x=0 \qquad [22]$$

where e and w represent the two ends of the control cell (**Figure 1(b)**). Introducing approximations for the flux at each ends and the source term in eqn [22] leads to

$$\frac{k_e(T_E-T_P)}{(\delta x)_e}-\frac{k_w(T_p-T_W)}{(\delta x)_w}+\bar{\gamma}\Delta x=0 \qquad [23]$$

where k_e and k_w are the heat conductivity at cell boundaries; T_E, T_P, and T_W are temperatures at nodal points; $\bar{\gamma}$ is the averaged source in the control cell; and $\Delta x=[(\delta x)_e+(\delta x)_w]/2$ is the size of the control cell (**Figure 1(b)**). Equation [23] is a discrete equation with nodal temperatures as unknown for the given control cell. Applying this procedure to all control cells leads to a system of equations with unknown temperatures at all grid points, similar to eqn [20] from an FD method.

FV methods have also been used extensively in modeling mantle convection, possibly starting with Ogawa et al. (1991). Tackley (1994) implemented an FV method coupled with a multigrid solver for 2-D/3-D Cartesian models. Ratcliff et al. (1996) developed a 3-D spherical shell model of mantle convection using an FV method. Harder and Hansen (2005), Stemmer et al. (2006), and Choblet et al. (2007) also developed FV mantle convection codes for spherical shell geometries using a cubed-sphere grid, while Kageyama and Sato (2004), Yoshida and Kageyama (2004), and Kameyama et al. (2008) built multigrid FV convection code on a 'yin–yang' grid. All these mantle convection studies used second-order accurate FV methods, although higher-order formulas can be formulated.

To solve fluid flow problems that are governed by the continuity and momentum equations, such as those in mantle convection, it is often convenient to use the pressure–velocity formulation, similar to FE method. Also, a staggered grid is often used in which pressures and velocities are defined at different locations of a control cell (Pantakar, 1980; Tackley, 1994; **Figure 1(c)–1(e)**). Such a staggered grid helps remove checkerboard pressure solutions. In the staggered grid, control cells for the momentum equation are different from those for the continuity equation. For example, for the continuity equation, a pressure node is at the center of a control cell, while velocities are defined at the cell boundaries such that each velocity component is perpendicular to the corresponding cell boundary (**Figure 1(e)**). For the momentum equation, the velocities are at the center of a cell, while the pressures are defined at cell boundaries (**Figure 1(c)** and **1(d)**).

For a given pressure field, applying the FV procedure leads to discrete equations for velocities. However, solutions to the velocity equations are only accurate if the pressure field is accurate. The continuity equation is used to correct the pressure field. This iterative procedure between the velocity and the pressure is implemented efficiently in a SIMPLE algorithm Pantakar (1980) that is used in the FV convection codes (Stemmer et al., 2006; Tackley, 1993, 1994). A variety of methods can be used to solve the discrete equations of velocities and pressure, similar to that in FD method. They may include successive overrelaxation and Gauss–Seidel iteration (Harder and Hansen, 2005; Stemmer et al., 2006), or a multigrid method (Tackley, 1994), or the alternating-direction implicit (i.e., ADI) method Monnereau and Yuen (2002).

Like in FE and FD methods, the FV method also requires special treatment of advection term $\vec{v}\cdot\nabla T$ in the energy equation. Either an upwind scheme or an iterative correction

scheme such as that proposed by Smolarkiewicz (1983, 1984) in Section 7.05.3.1 on FD can be used in the FV method to treat the advection term.

7.05.3.3 Spectral Methods

The spectral method is a classical method, which is motivated by its analytic popularity and inherent accuracy. It is based on the concept of orthogonal eigenfunction expansion based on the differential operators associated with the Laplacian equation and works on orthogonal curvilinear coordinate systems. For mantle convection problems, this means that we can express the horizontal dependence by using Fourier expansion for Cartesian geometry and spherical harmonics for the 3-D spherical shell. Symbolically, we can write down this expansion as

$$F(x_1, x_2, x_3) = \sum_{ijk} a_{ijk} f(x_3) g(x_1) h(x_2) \qquad [24]$$

where F is the field variable being expanded, a_{ijk} is the spectral coefficient, x_1 and x_2 are the horizontal coordinates, x_3 is the vertical or radial coordinate, g and h are the eigenfunctions being employed, and f is a function describing the vertical or radial dependence. We note that f can be determined by solving a two-point boundary value, using propagator matrices, or an orthogonal function expansion, as in the case of the Chebyshev polynomials, or using the FD (e.g., Cserepes and Rabinowicz, 1985; Cserepes et al., 1988; Gable et al., 1991; Glatzmaier, 1988; Machetel et al., 1986; Travis et al., 1990).

Balachandar and Yuen (1994) developed a spectral-transform 3-D Cartesian code for mantle convection, based on expansion of the Chebyshev functions for solving the two-point boundary value problem in the vertical direction and fast Fourier transforms along the two horizontal directions. This method has also been extended to variable viscosity problems with the help of the Krylov subspace iterative technique (Saad and Schultz, 1986) for solving iteratively the momentum equation and has gone up to viscosity contrast of 1000. The first 3-D codes in spherical shell convection were developed by Machetel et al. (1986) who used an FD scheme in the radial direction and spherical harmonic decomposition over the longitudinal and latitudinal directions and by Glatzmaier (1988) who employed the Chebyshev polynomials in solving the radial direction and spherical harmonics for the spherical surface. The Chebyshev polynomials allow for very high accuracy near the thermal boundary layers. They can also be expanded about internal boundary layers, such as the one at 670 km discontinuity. Zhang and Yuen (1995, 1996a,b) had devised a 3-D code for spherical geometry based on higher-order FD method (Fornberg, 1995) and spectral expansion in the field variables in spherical harmonics along the circumferential directions. They also developed an iterative technique for the momentum equation to solve for variable viscosity with a lateral viscosity contrast up to 200 (Zhang and Yuen, 1996b) in compressible convection. It is important to check also the spectra of the viscosity decay with the degree of the spherical harmonics (Zhang and Yuen, 1996a) in order to ascertain whether the Gibbs phenomenon is present.

Spectral methods are very accurate, much more so than FEs and FDs for the same computational efforts, and their efficacy in terms of convergence can be assessed by examining the rate of decay of the energy in the spectrum with increasing number of terms in the spectral expansion in eqn [24] (Peyret and Taylor, 1983). Spectral methods also have a couple of other distinct advantages. First, they are very easy to implement, because they reduce the problems to either an algebraic set or simple weakly coupled ordinary differential equations, as discussed earlier. Second, using fast transform algorithms such as fast Fourier transform, spectral methods can be very efficient and fast on a single-processor computer. They have been extended to parallel processors, using libraries.

However, spectral methods only work well for constant material properties or depth-dependent properties in simple geometries (Balachandar and Yuen, 1994). Furthermore, spectral techniques do not go as far in terms of viscosity contrasts ($<$ a factor of 200) for variable viscosity convection (Balachandar et al., 1995; Zhang and Yuen, 1996b). We note, however, that Schmalholz et al. (2001) showed that in a folding problem, their spectral method could go up to a viscosity contrast of around ten thousand. Perhaps, a better choice of a preconditioner in the solution of the momentum equation in the spectral expansion of a variable viscosity problem will enable a larger viscosity contrast than 10^4. This issue is still open and remains a viable research topic. Spectral methods are also difficult to be used efficiently on parallel computers because of the global basis functions used in these methods. At present, spectral methods, although elegant mathematically, are not as effective as other numerical methods for solving realistic mantle convection problems. Spectral methods still remain the main driver for geodynamo problems (e.g., Glatzmaier and Roberts, 1995; Kuang and Bloxham, 1999) because of the exclusive choice of constant material properties in solving the set of magneto-hydrodynamic equations. Lastly, we wish to mention briefly a related spectral method called the spectral-element method (Patera, 1984; Maday and Patera, 1989), which is now being used in geophysics in seismic wave propagation (Komatitsch and Tromp, 1999). This approach may hold some promise for variable viscosity mantle convection.

7.05.4 An FE Method

FE methods are effective in solving differential equations with complicated geometry and material properties. FE methods have been widely used in the studies of mantle dynamics (Baumgardner, 1985; Bunge et al., 1997; Christensen, 1984; King et al., 1990; Moresi and Gurnis, 1996; van den Berg et al., 1993; Zhong et al., 2000, 2008). This section will go through some of the basic steps in using FE methods in solving governing equations for thermal convection.

The FE formulation for Stokes' flow that is described by eqns [1] and [2] is independent from that for the energy equation. Hughes (2000) gave detailed description on a Galerkin weak form FE formulation for Stokes' flow. Brooks (1981) developed a streamline upwind Petrov–Galerkin formulation (SUPG) for the energy equation involving advection and diffusion. These two formulations remain popular for solving these types of problems (Hughes, 2000) and are employed in mantle convection codes ConMan (King et al., 1990) and

Citcom/CitcomS (Moresi and Gurnis, 1996; Zhong et al., 2000, 2008). The descriptions presented here are tailored from Brooks (1981), Hughes (2000), and Ramage and Wathen (1994) specifically for thermal convection in an incompressible medium, and they are also closely related to codes ConMan and Citcom.

7.05.4.1 Stokes' Flow: A Weak Formulation, Its FE Implementation and Solution

7.05.4.1.1 A weak formulation

The Galerkin weak formulation for Stokes' flow can be stated as follows: find the flow velocity $u_i = v_i + g_i$ and pressure P, where g_i is the prescribed boundary velocity from eqn [6] and $v_i \in \mathcal{V}$, and $P \in \mathcal{P}$, where $\mathcal{V}$ is a set of functions in which each function, w_i, is equal to zero on Γ_{g_i} and $\mathcal{P}$ is a set of functions q, such that for all $w_i \in \mathcal{V}$ and $q \in \mathcal{P}$:

$$\int_\Omega w_{i,j}\sigma_{ij}\mathrm{d}\Omega - \int_\Omega q u_{i,i}\mathrm{d}\Omega = \int_\Omega w_i f_i \mathrm{d}\Omega + \sum_{i=1}^{n_{sd}} \int_{\Gamma_{h_i}} w_i h_i \mathrm{d}\Gamma \qquad [25]$$

w_i and q are also called weighting functions. Equation [25] is equivalent to eqns [1] and [2] and boundary conditions eqn [6], provided that $f_i = RaT\delta_{iz}$ (Hughes, 2000). Equation [25] can be written as

$$\begin{aligned}&\int_\Omega w_{i,j}c_{ijkl}v_{k,l}\mathrm{d}\Omega - \int_\Omega q v_{i,i}\mathrm{d}\Omega - \int_\Omega w_{i,i}P\mathrm{d}\Omega \\ &= \int_\Omega w_i f_i \mathrm{d}\Omega + \sum_{i=1}^{n_{sd}} \int_{\Gamma_{h_i}} w_i h_i \mathrm{d}\Gamma - \int_\Omega w_{i,j}c_{ijkl}g_{k,l}\mathrm{d}\Omega\end{aligned} \qquad [26]$$

where

$$c_{ijkl} = \eta\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) \qquad [27]$$

is derived from constitutive eqn [4].

It is often convenient to rewrite

$$w_{i,j}c_{ijkl}v_{k,l} = \varepsilon\left(\vec{w}\right)^T D\varepsilon\left(\vec{v}\right) \qquad [28]$$

where for 2-D plane strain problems,

$$\varepsilon\left(\vec{v}\right) = \begin{Bmatrix} v_{1,1} \\ v_{2,2} \\ v_{1,2}+v_{2,1} \end{Bmatrix}, D = \begin{bmatrix} 2\eta & 0 & 0 \\ 0 & 2\eta & 0 \\ 0 & 0 & \eta \end{bmatrix} \qquad [29]$$

(note here the change in the definition of the off-diagonal components of the strain-rate tensor). It is straightforward to write the expressions for other coordinate systems including 3-D Cartesian, axisymmetric (Hughes, 2000), or spherical geometry (Zhong et al., 2000, 2008).

Suppose that we discretize the solution domain Ω using a set of grid points (Figure 2) so that the velocity and pressure fields and their weighting functions can be expressed anywhere in the domain in terms of their values at the grid (nodal) points and the so-called *shape functions* that interpolate the grid points:

$$\begin{aligned}&\vec{v} = v_i\vec{e}_i = \sum_{A\in\Omega^v-\Gamma^v_{g_i}} N_A v_{iA}\vec{e}_i,\ \vec{w} = w_i\vec{e}_i = \sum_{A\in\Omega^v-\Gamma^v_{g_i}} N_A w_{iA}\vec{e}_i, \\ &\vec{g} = \sum_{A\in\Gamma^v_{g_i}} N_A g_{iA}\vec{e}_i\end{aligned} \qquad [30]$$

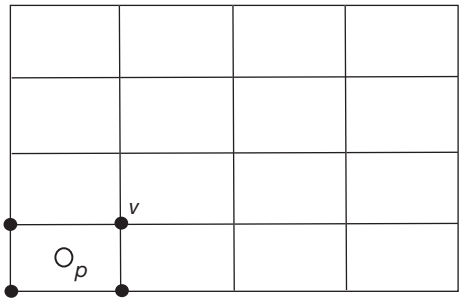


Figure 2 Finite-element discretization and grid in 2-D. For a 4-node bilinear element, the pressure is defined at the center of the element, while velocities are defined at the four corners.

$$P = \sum_{B\in\Omega^p} M_B P_B, q = \sum_{B\in\Omega^p} M_B q_B \qquad [31]$$

where N_A is the shape functions for velocity at node A, M_B is the shape functions for pressure at node B, Ω^v is the set of velocity nodes, Ω^p is the set of pressure nodes, and $\Gamma^v_{g_i}$ is the set of velocity nodes along boundary Γ_{g_i}. Note that the velocity shape functions and velocity nodes can be (and usually are) different from those for pressure (Figure 2).

Substituting eqn [30] into [28] leads to the following equation:

$$\begin{aligned}\varepsilon\left(\vec{w}\right)^T D\varepsilon(\vec{v}) &= \varepsilon\left(\sum_{A\in\Omega^v-\Gamma^v_{g_i}} N_A w_{iA}\vec{e}_i\right)^T D\varepsilon\left(\sum_{B\in\Omega^v-\Gamma^v_{g_j}} N_B v_{jB}\vec{e}_j\right) \\ &= \left[\sum_{A\in\Omega^v-\Gamma^v_{g_i}} \varepsilon\left(N_A\vec{e}_i\right)^T w_{iA}\right] D\left[\sum_{B\in\Omega^v-\Gamma^v_{g_j}} \varepsilon\left(N_B\vec{e}_j\right)v_{jB}\right] \\ &= \sum_{A\in\Omega^v-\Gamma^v_{g_i}} w_{iA}\left[\sum_{B\in\Omega^v-\Gamma^v_{g_j}} \vec{e}_i{}^T B_A^T D B_B \vec{e}_j v_{jB}\right]\end{aligned} \qquad [32]$$

where for 2-D plane strain problems,

$$B_A = \begin{bmatrix} N_{A,1} & 0 \\ 0 & N_{A,2} \\ N_{A,2} & N_{A,1} \end{bmatrix} \qquad [33]$$

Substituting eqns [30] and [31] into [26] leads to the following equation:

$$\begin{aligned}&\sum_{A\in\Omega^v-\Gamma^v_{g_i}} w_{iA}\left[\sum_{B\in\Omega^v-\Gamma^v_{g_j}}\left(\vec{e}_i{}^T\int_\Omega B_A{}^T D B_B \mathrm{d}\Omega\vec{e}_j v_{jB}\right)\right. \\ &\qquad \left. -\sum_{B\in\Omega^p}\left(\vec{e}_i\int_\Omega N_{A,i}M_B\mathrm{d}\Omega P_B\right)\right] \\ &\qquad -\sum_{A\in\Omega^p}\left[q_A\sum_{B\in\Omega^v-\Gamma^v_{g_j}}\left(\int_\Omega M_A N_{B,j}\mathrm{d}\Omega\vec{e}_j v_{jB}\right)\right] \\ &\qquad = \sum_{A\in\Omega^v-\Gamma^v_{g_i}} w_{iA}\left[\int_\Omega N_A\vec{e}_i f_i\mathrm{d}\Omega + \sum_{i=1}^{n_{sd}}\int_{\Gamma_{h_i}} N_A\vec{e}_i h_i\mathrm{d}\Gamma\right. \\ &\qquad \left. -\sum_{B\in\Gamma^v_{g_j}}\left(\vec{e}_i{}^T\int_\Omega B_A{}^T D B_B\mathrm{d}\Omega\vec{e}_j g_{jB}\right)\right]\end{aligned} \qquad [34]$$

Because eqn [34] holds for any weighting functions w_{iA} and q_A, it implies the following two equations:

$$\sum_{B\in\Omega^v-\Gamma^v_{s_j}}\left(\vec{e}_i^{\,T}\int_\Omega {B_A}^T D B_B \mathrm{d}\Omega \vec{e}_j v_{jB}\right)-\sum_{B\in\Omega^p}\left(\vec{e}_i\int_\Omega N_{A,i}M_B\mathrm{d}\Omega P_B\right)$$
$$=\int_\Omega N_A\vec{e}_i f_i\mathrm{d}\Omega+\sum_{i=1}^{n_{sd}}\int_{\Gamma_{h_i}}N_A\vec{e}_i h_i\mathrm{d}\Gamma-\sum_{B\in\Gamma^v_{s_j}}\left(\vec{e}_i^{\,T}\int_\Omega {B_A}^T D B_B\mathrm{d}\Omega\vec{e}_j g_{jB}\right) \quad [35]$$

$$\sum_{B\in\Omega^v-\Gamma^v_{s_j}}\left(\int_\Omega M_A N_{B,j}\mathrm{d}\Omega\vec{e}_j v_{jB}\right)=0 \quad [36]$$

Combining eqns [35] and [36] into a matrix form leads to

$$\begin{bmatrix}K & G\\ G^T & 0\end{bmatrix}\begin{Bmatrix}V\\P\end{Bmatrix}=\begin{Bmatrix}F\\0\end{Bmatrix} \quad [37]$$

where the vector V contains the velocity at all the nodal points, the vector P is the pressure at all the pressure nodes, the vector F is the total force term resulting from the three terms on the right-hand side of eqns [35] or [36], and the matrices K, G, and G^T are the stiffness matrix, discrete gradient operator, and discrete divergence operator, respectively, which are derived from the first and second terms of eqns [35] and [36], respectively. Specifically, the stiffness matrix is given by

$$K_{lm}=\vec{e}_i^{\,T}\int_\Omega {B_A}^T D B_B\mathrm{d}\Omega\vec{e}_j \quad [38]$$

where subscripts A and B are the global velocity node numbers as in eqn [30], i and j are the degree of freedom numbers ranging from 1 to n_{sd}, and l and m are the global equation numbers for the velocity ranging from 1 to $n_v n_{sd}$ where n_v is the number of velocity nodes.

7.05.4.1.2 An FE implementation

We now present an FE implementation of the Galerkin weak formulation for Stokes' flow and the resulting expressions of different terms in [37]. A key point of the FE approach is that all of the equations to be solved are written in the form of integrals over the solution domain and can, therefore, without approximation, be written as a sum of integrals over convenient subdomains, and the matrix eqn [37] may be decomposed into overlapping sums of contributions from these domains.

Let us first introduce the elements and shape functions. A key feature of the standard FE method is that a local basis function or shape function is used such that the value of a variable within an element depends only on that at nodal points of the element. The shape functions are generally chosen such that they have the value of unity on their parent node and zero at all other nodes and zero outside the boundary of the element. Unless there is a special, known form to the solution, it is usual to choose simple polynomial shape functions and form their products for additional dimensions. In addition, this interpolating requirement constrains the patterns of nodes in an element and which elements can be placed adjacent to each other. Bathe (1996) gave an excellent overview of the issues. (As usual, there are some useful exceptions to this rule, one being the element-free Galerkin method where smooth, overlapping interpolating kernels are used. This does mean, however, that the global problem cannot be trivially decomposed into local elements.) Zienkiewicz et al. (2005) is also an excellent source of reference for these topics.

For simplicity, we consider a 2-D domain with quadrilateral elements. We employ mixed elements in which there are four velocity nodes per element, each of which occupies a corner of the element, while the only pressure node is at the center of the element (Figure 2). For these quadrilateral elements, the velocity interpolation in each element uses bilinear shape functions, while the pressure is constant for each element.

As a general remark on FE modeling of deformation/flow of incompressible media, it is important to keep interpolation functions (shape functions) for velocities at least one order higher than those for pressure, as we did for our quadrilateral elements (Hughes, 2000). Spurious flow solutions may arise sometimes even if this condition is satisfied. The best-known example is the 'mesh locking' that arises from linear triangle elements with constant pressure per element for which incompressibility (i.e., a fixed elemental area) constraint per element demands zero deformation/flow everywhere in the domain (Hughes, 2000).

For any given element e, velocity and pressure within this element can be expressed through the following interpolation:

$$\vec{v}=v_i\vec{e}_i=\sum_{a=1}^{n_{en}}N_a v_{ia}\vec{e}_i,\ \vec{w}=w_i\vec{e}_i=\sum_{b=1}^{n_{en}}N_b w_{ib}\vec{e}_i,\ \vec{g}=g_i\vec{e}_i$$
$$=\sum_{a=1}^{n_{en}}N_a g_{ia}\vec{e}_i \quad [39]$$

$$P=\sum_{a=1}^{n_{ep}}M_a P_a, q=\sum_{a=1}^{n_{ep}}M_a q_a \quad [40]$$

where n_{en} and n_{ep} are the numbers of velocity and pressure nodes per element, respectively, and $n_{en}=4$ and $n_{ep}=1$ for our quadrilateral elements. The shape function N_a for $a=1,\ldots,n_{en}$, depends on coordinates, and N_a is 1 at node a and linearly decreases to zero at other nodes of the element. The locality of the shape functions greatly simplifies implementation and computational aspects of the Galerkin weak formulation. For example, the integrals in eqns [35] and [36] may be decomposed into sum of integrals from each element, and the matrix eqn [37] may be decomposed into sums of elemental contributions. Specifically, we may now introduce elemental stiffness matrix, discrete gradient and divergence operators, and force term:

$$k^e=\left[k^e_{lm}\right],\ g^e=\left[g^e_{\mathrm{ln}}\right],\ f^e=\left\{f^e_l\right\} \quad [41]$$

where $1\le l,m\le n_{en}n_{sd}$, $1\le n\le n_{ep}$ (note that for quadrilateral elements, $n_{en}=4$, $n_{ep}=1$, and $n_{sd}=2$), k^e is a square matrix of $n_{en}n_{sd}$ by $n_{en}n_{sd}$, and g^e is a matrix of $n_{en}n_{sd}$ by n_{ep},

$$k^e_{lm}=\vec{e}_i^{\,T}\int_{\Omega^e}{B_a}^T D B_b\mathrm{d}\Omega\vec{e}_j \quad [42]$$

where $l=n_{sd}(a-1)+i$, $m=n_{sd}(b-1)+j$, $a,b=1,\ \ldots,\ n_{en}$, and $i,j=1,\ldots,n_{sd}$,

$$g^e_{\mathrm{ln}}=-\vec{e}_i\int_{\Omega^e}N_{a,i}M_n\mathrm{d}\Omega \quad [43]$$

where $n=1,\ \ldots,\ n_{ep}$, and the rest symbols have the same definitions as before,

$$f_l^e = \int_{\Omega^e} N_a f_i \mathrm{d}\Omega + \int_{\Gamma^e_{h_i}} N_a h_i \mathrm{d}\Gamma - \sum_{m=1}^{n_{sd}n_{en}} k^e_{lm} g^e_m \qquad [44]$$

Determinations of these elemental matrices and force term require evaluations of integrals over each element with integrands that involve the shape functions and their derivatives. It is often convenient to use isoparametric elements for which the coordinates and velocities in an element have the interpolation schemes (Hughes, 2000). For example, for 2-D quadrilateral elements that we discussed earlier, the velocity shape function for node a of an element in a parent domain with coordinates $\xi \in (-1, 1)$ and $\eta \in (-1, 1)$ is given as

$$N_a(\xi, \eta) = 1/4(1 + \xi_a\xi)(1 + \eta_a\eta) \qquad [45]$$

where (ξ_a, η_a) is $(-1,-1)$, $(1,-1)$, $(1,1)$, and $(-1,1)$ for $a = 1$, 2, 3, and 4, respectively. The pressure shape function for M_a is 1, as there is only one pressure node per element. Although the integrations in [42]–[44] are in the physical domain (i.e., x_1 and x_2 coordinates) rather than the parent domain, they can be expressed in the parent domain through a coordinate transformation.

In practice, these element integrals are calculated numerically by some form of quadrature rule. In 2-D, Gaussian quadrature rules are optimal and are usually recommended; in 3-D, other rules may be more efficient but are not commonly used for reasons of programming simplicity. (Hughes (2000) documented integration procedure in detail.) There are a small number of cases where it may be worthwhile using a nonstandard procedure to integrate an element. The most common is where the constitutive parameters (i.e., D in eqn [29]) change within the element; a higher-order quadrature scheme than the standard recommended one can give improved accuracy in computing k^e. When the constitutive parameters are strongly history-dependent, D is known only where there are Lagrangian sample points; if these are used directly to integrate k^e, the Lagrangian integration point methods (such as the Material Point Method (MPM)) result (Moresi et al., 2003; Sulsky et al., 1994 for application in geodynamics).

With elemental k^e, g^e, and f^e determined, it is straightforward to assemble them into global matrix eqn [37]. If an iterative solution method is used to solve [37], one may carry out calculations of the left-hand side of [37] element by element without assembling the global matrix equation form [37].

The boundary conditions described in eqn [44] are simple to implement when the boundary is aligned with the coordinate system and when the boundary condition is purely velocity or purely boundary tractions. A difficulty arises in the case of boundaries that are not aligned with the coordinate system and that specify a mixture of boundary tractions and velocities. This form of boundary condition is common in engineering applications of FEM, designing mechanical components, for example, and in geodynamics, it occurs when modeling arbitrarily oriented contact surfaces such as faults and a spherical boundary in a Cartesian solution domain. Unless treated appropriately, this form of boundary condition leads to a constraint on a linear combination of the degrees of freedom at each constrained node. The solution is straightforward: instead of using the global coordinate system to form the matrix equation, we first transform the equations at any affected mesh point to a local coordinate system that does conform to the boundary geometry (Desai and Abel, 1972).

Let R be a rotation matrix that transforms the global coordinate system $XYZO$ to a local one $X'Y'Z'O'$ aligned with the boundary at node n. The transformed stiffness matrix and force vector contributions for this node then become

$$K_n^{`} = R_n^T K_n R_n \qquad [46]$$

and

$$F_n^{`} = R_n^T F_n \qquad [47]$$

R can be formulated for each node or it may be assembled for an element or for the global matrices, but, other than the nodes where the 'skew' boundary conditions occur, the block entries are simply the identity matrix. It is possible to use this procedure at every node in a system with a natural coordinate system (e.g., the spherical domain) so that one can switch freely between a spherical description of forces and velocities where convenient and an underlying Cartesian formulation of the constitutive relations. This method was used in modeling faults in Zhong et al. (1998).

7.05.4.1.3 The Uzawa algorithm for the matrix equation

Similar to FD or FV methods, the FE discretization of the differential equations leads to a matrix equation such as [37]. The remaining question is to solve the matrix equation, which we now discuss in this subsection. We can use either direct or iterative methods to solve the matrix equation, depending on problems that we are interested, similar to what we discussed earlier for matrix eqn [20] from FD method. Here, we will focus on iterative solution methods because they require significantly less memory and computation than direct solution approaches. Iterative solution approaches are the only feasible and practical approaches for large 3-D problems. Later, we will briefly discuss a penalty formulation for the incompressible Stokes' flow that requires a direct solution approach and is therefore only effective for 2-D problems.

This system of equation, with a zero block on the diagonal since pressure does not couple to itself, is referred to as a saddle point system. This indicates that it is indefinite, having both positive and negative eigenvalues, making it unsuitable for the conjugate gradient iteration and many simple preconditioners. It is also difficult to factor explicitly since pivot elements in LU cannot be zero. However, we can divide the system into separate problems for pressure and velocity, in order to take advantage of the structure of both. The matrix K is symmetrical and positive definite with a constant viscosity, although this will often hold even for temperature- and stress-dependent viscosity. We can use this to our advantage in finding a solution strategy. An efficient method is the Uzawa algorithm, which is implemented in Citcom code (Moresi and Solomatov, 1995). In the Uzawa algorithm, the matrix eqn [37] is broken into two coupled systems of equations (Atanga and Silvester, 1992; Ramage and Wathen, 1994):

$$KV + GP = F \qquad [48]$$

$$G^T V = 0 \qquad [49]$$

Combining these two equations and eliminating V to form the Schur complement system for pressure (Hughes, 2000),

$$(G^T K^{-1} G)P = G^T K^{-1} F \qquad [50]$$

Notice that matrix $\hat{K} = G^T K^{-1} G$ is symmetrical positive definite. Although in practice eqn [50] cannot be solved directly for P due to difficulties in obtaining K^{-1}, the Krylov subspace iterative solution methods can be used as they do not require the explicit construction of $\hat{K}$, only products of $\hat{K}$ with a sequence of vectors. As an example, we consider the application of the conjugate gradient algorithm to [50] as first described by Ramage and Wathen (1994).

With the conjugate gradient algorithm, for symmetrical positive definite $\hat{K}$, the solution to a linear system of equations $\hat{K}P = \hat{F}$ can be obtained with the operations in **Figure 3(a)** (Golub and van Loan, 1989, p. 523). For eqns [48] and [50] for both velocities and pressure, with initial guess pressure $P_0 = 0$, the initial velocity V_0 can be obtained from

$$KV_0 = F \quad \text{or} \quad V_0 = K^{-1} F \qquad [51]$$

and the initial residual for pressure eqn [50], r_0, and search direction, s_1, are $r_0 = s_1 = \hat{F} = G^T K^{-1} F = G^T V_0$ (**Figure 3(a)**). To determine the search step α_k in the conjugate gradient algorithm, we need to compute the product of search direction s_k with $\hat{K}$, $s_k^T \hat{K} s_k$ (**Figure 3**). This product can be evaluated without explicitly constructing $\hat{K}$ for the following reasons.

The product can be written as

$$s_k^T \hat{K} s_k = s_k^T G^T K^{-1} G s_k = (G s_k)^T K^{-1} G s_k \qquad [52]$$

If we define u_k, such that

$$K u_k = G s_k \quad \text{or} \quad u_k = K^{-1} G s_k \qquad [53]$$

then we have

$$s_k^T \hat{K} s_k = (G s_k)^T K^{-1} G s_k = (G s_k)^T u_k \qquad [54]$$

This indicates that if we solve [53] for u_k with Gs_k as the force term, the product $s_k^T \hat{K} s_k$ can be obtained without actually forming $\hat{K}$. Similarly, $\hat{K} s_k$ in updating the residual r_k (**Figure 3(a)**) can be obtained without forming $\hat{K}$, because $\hat{K} s_k = G^T K^{-1} G s_k = G^T u_k$.

As the pressure P is updated via $P_k = P_{k-1} + \alpha_k s_k$ from the conjugate gradient algorithm (**Figure 3**), the velocity field can also be updated accordingly via

$k = 0$; $P_0 = 0$; $r_0 = \hat{F}$
while $|r_k| > \in$(a given tolerance):
  $k = k + 1$
  if $k = 1$:
    $s_1 = r_0$
  else:
    $\beta_k = r_{k-1}^T r_{k-1} / r_{k-2}^T r_{k-2}$
    $s_k = r_{k-1} + \beta_k s_{k-1}$
  end
  $\alpha_k = r_{k-1}^T r_{k-1} / s_k^T \hat{K} s_k$
  $P_k = P_{k-1} + \alpha_k s_k$
  $r_k = r_{k-1} - \alpha_k \hat{K} s_k$
end
$P = P_k$

(a)

$k = 0$; $P_0 = 0$; $r_0 = \hat{F}$
while $|r_k| > \in$(a given tolerance):
  Solve $Mz_k = r_k$
  $k = k + 1$
  if $k = 1$:
    $s_1 = z_0$
  else:
    $\beta_k = r_{k-1}^T z_{k-1} / r_{k-2}^T z_{k-2}$
    $s_k = r_{k-1} + \beta_k s_{k-1}$
  end
  $\alpha_k = r_{k-1}^T z_{k-1} / s_k^T \hat{K} s_k$
  $P_k = P_{k-1} + \alpha_k s_k$
  $r_k = r_{k-1} - \alpha_k \hat{K} s_k$
end
$P = P_k$

(b)

$k = 0$; $P_0 = 0$
Solve $KV_0 = F$
$r_0 = \hat{F} = G^T K^{-1} F = G^T V_0$
while $|r_k| > \in$(a given tolerance):
  $k = k + 1$
  if $k = 1$:
    $s_1 = r_0$
  else:
    $\beta_k = r_{k-1}^T r_{k-1} / r_{k-2}^T r_{k-2}$
    $s_k = r_{k-1} + \beta_k s_{k-1}$
  end
  Solve $Ku_k = Gs_k$
  $\hat{K} s_k = G^T K^{-1} s_k = G^T u_k$
  $\alpha_k = r_{k-1}^T r_{k-1} / s_k^T \hat{K} s_k = r_{k-1}^T r_{k-1} / s_k^T G^T u_k$
  $P_k = P_{k-1} + \alpha_k s_k$
  $r_k = r_{k-1} - \alpha_k \hat{K} s_k = r_{k-1} - \alpha_k G^T u_k$
end
$P = P_k$
Solve $KV = F - GP$

(c)

$k = 0$; $P_0 = 0$
Solve $KV_0 = F$
$r_0 = \hat{F} = G^T K^{-1} F = G^T V_0$
while $|r_k| > \in$(a given tolerance):
  Solve $Mz_k = r_k$
  $k = k + 1$
  if $k = 1$:
    $s_1 = z_0$
  else:
    $\beta_k = r_{k-1}^T z_{k-1} / r_{k-2}^T z_{k-2}$
    $s_k = z_{k-1} + \beta_k s_{k-1}$
  end
  Solve $Ku_k = Gs_k$
  $\hat{K} s_k = G^T K^{-1} s_k = G^T u_k$
  $\alpha_k = r_{k-1}^T z_{k-1} / s_k^T \hat{K} s_k = r_{k-1}^T z_{k-1} / s_k^T G^T u_k$
  $P_k = P_{k-1} + \alpha_k s_k$
  $r_k = r_{k-1} - \alpha_k \hat{K} s_k$
end
$P = P_k$
Solve $KV = F - GP$

(d)

Figure 3 Variants of the conjugate gradient algorithm for solving the matrix problem $\hat{K}P = \hat{F}$ for the unknown vector P. (a) is the classic conjugate gradient algorithm given by Golub and van Loan (1989); (b) is the preconditioned form. In (c), we show how to formulate the algorithm for the case where $\hat{K} = G^T K^{-1} G$ and $\hat{F} = G^T K^{-1} F$, which arise from the transformation of the saddle point problem, and (d) is the preconditioned form of this version.

$$V_k = V_{k-1} - \alpha_k u_k \quad [55]$$

This can be seen from the following derivation. At iteration step $k-1$, the pressure and velocity are P_{k-1} and V_{k-1}, respectively, and they satisfy eqn [48]

$$KV_{k-1} + GP_{k-1} = F \quad [56]$$

At iteration step k, the updated pressure is P_k, and the velocity $V_k = V_{k-1} + v$, where v is the unknown increment to be determined. Substituting P_k and V_k into [48] and considering $P_k = P_{k-1} + \alpha_k s_k$, $V_k = V_{k-1} + v$, and [54] lead to

$$Kv + \alpha_k G s_k = 0, \quad \text{or} \quad v = -\alpha_k K^{-1} G s_k \quad [57]$$

From [53], it is clear that the velocity increment $v = -\alpha_k u_k$, and consequently, eqn [55] updates the velocity.

The full algorithm is given in **Figure 3(c)** and follows Ramage and Wathen (1994). The changes required to implement this approach for the conjugate gradient are minor (compare **Figure 3(a)** and **3(c)**) and can be entirely abstracted to make this approach general for any other Krylov subspace methods such as biconjugate gradient, GMRES, and flexible GMRES.

A key ingredient for this approach to be efficient is a reliable and efficient solver for eqn [53] since this is required once to calculate the residual, once per iteration, and once if the backsolve for V is required. There is no prior guess for the intermediate velocity in [53] and the right-hand side of this equation does not systematically reduce as the outer iteration converges. Furthermore, although there has been some work on using inexact solutions of [53] to reduce the cost of each iteration, the fact that the solution is cumulative in P (and in V) for most Krylov methods, including conjugate gradient, means that errors are neither corrected nor detected in the stopping condition (a good introductory discussion to this subject is given by Eshof et al., 2005). In practice, we find that it is most reliable to set an absolute tolerance on the residual of [53] of at least one order of magnitude more stringent than that of the required residual to [50].

The efficiency of the Uzawa algorithm can be significantly improved with preconditioners for $\hat{K}$, as to be discussed in **Section 7.05.6.3**. It should be also pointed out that the Uzawa algorithm outlined earlier can also be used in connection with other numerical methods including FV and FD methods, provided that the matrix equations for pressure and velocities from these numerical methods have the form of [37].

7.05.4.1.4 Multigrid solution strategies

The stiffness matrix K is symmetrical positive definite, and this allows for numerous possible solution approaches. For example, multigrid solvers have been used for solving eqn [53] (Moresi and Solomatov, 1995) as in Citcom code. Newer versions of Citcom including CitcomS/CitcomCU employ full multigrid solvers with a consistent projection scheme for eqn [53] and are even more efficient (Zhong et al., 2000).

The multigrid method works by formulating the FE problem on a number of different scales – usually a set of grids that are nested one within the other sharing common nodes (see **Figure 4(a)**), similar to how multigrid methods are used in FD and FV methods. The solution progresses on all of the grids at the same time with each grid, eliminating errors at a different scale. The effect is to propagate information very rapidly between different nodes in the grid that would otherwise be prevented by the local support of the element shape functions. In fact, by a single traverse from fine to coarse grid and back, all nodes in the mesh can be directly connected to every other – allowing nodes that are physically coupled but remote in the mesh to communicate directly during each iteration cycle. This matches the physical structure of Stokes' flow problem in which stresses are transmitted instantaneously to all parts of the system in response to changes anywhere in the buoyancy forces or boundary conditions.

The multigrid effect relies upon using an iterative solver on each of many nested grid resolutions that acts like a smoother on the residual error at the characteristic scale of that particular grid. Gauss–Seidel iteration is a very common choice because it has exactly this property, although its effectiveness depends on the order in which degrees of freedom are visited by the solver, and, consequently, it can be difficult to implement in parallel. On the coarsest grid, it is possible to use a direct solver because the number of elements is usually very small.

For an elliptical operator such as the stiffness matrix, K, of Stokes' flow problem in eqn [53], we write

$$K_h v_h = F_h \quad [58]$$

where the h subscript indicates that the problem has been discretized to a mesh of fineness h. An initial estimate of the velocity on grid h, v_h, can be improved by a correction Δv_h determining the solution to

$$K_h \Delta v_h = F_h - K_h v_h \quad [59]$$

Suppose we obtain our approximate initial estimate by solving the problem on a coarse grid. The reduction of the

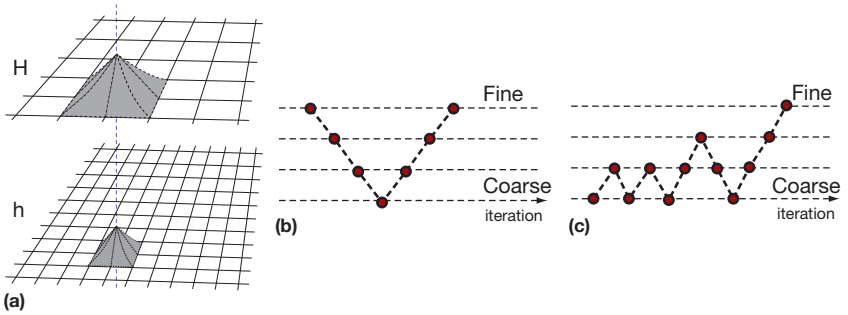


Figure 4 A nested grids with fine and coarse meshes (a), structure of V-cycle multigrid iteration (b), and structure of a full multigrid iteration (c).

number of degrees of freedom leads to a more manageable problem that can be solved quickly. The correction term is therefore

$$K_h \Delta v_h = F_h - K_h R_h^H v_H \quad [60]$$

where H indicates a coarser level of discretization and the operator R_h^H is an interpolation from the coarse level H to the fine level h.

To find v_H, we need to solve a coarse approximation to the problem

$$K_H v_H = F_H \quad [61]$$

where K_H and F_H are the coarse-level equivalent of the stiffness matrix and force vector. One obvious way to define these is to construct them from a coarse representation of the problem on the mesh H exactly as would be done on h. An alternative is to define

$$K_H = R_H^h K_h R_h^H \text{ and } F_H = R_H^h F_h \quad [62]$$

where R_h^H is a 'restriction' operator that has the opposite effect to the interpolation operation in that it lumps nodal contributions from the fine mesh onto the coarse mesh.

The power of the algorithm comes from a recursive application. The coarse grid correction is also calculated through the use of a still-coarser grid and so on, until the problem is so small that an exact solution can be obtained very rapidly. One very simple, but instructive, algorithm for hierarchical residual reduction is the sawtooth cycle given in Figure 5, and its logical layout is same as the V cycle in Figure 4(b).

The step in which the velocity correction is 'improved' is an iterative method for reducing the residual at the current level that has the property of smoothing the error strongly at the current mesh scale. At each level, these smoothing operators reduce the residual most strongly on the scale of the discretization – the hierarchical nesting of different mesh sizes allows the residual to be reduced at each scale very efficiently (see Yavneh, 2006 and Briggs et al. 2000 for a more lengthy discussion).

The projection and interpolation operators have to be chosen fairly carefully to avoid poor approximations to the problem at the coarse levels and ineffectual corrections propagated to the fine levels. The interpolation operator is defined naturally from the shape functions at the coarse levels. The projection operator is then defined to complement this choice (the operators should be the adjoint). The projection operator for the velocity and residual can differ from that for the viscosity, and this can be optimized to yield significantly better convergence (see Jadamec et al., 2012).

The sawtooth cycle outlined in this section is the simplest multigrid algorithm. Developments include improving the residual at each level of the restriction as well as the interpolation, known as a 'V cycle,' and cycles in which the residual is interpolated only part way through the hierarchy before being reprojected and subjected to another set of improvements (a 'W cycle').

The full multigrid algorithm introduces a further level of complexity. Instead of simply casting the problem at a single level and projecting/improving the residual on a number of grids, the whole problem is defined for all the grids. In this way, the initial fine-grid approximation is obtained by interpolating from the solution to the coarsest-grid problem. The solution at each level is still obtained by projecting to the finest level and reducing the residual at each projection step. The resulting cycle is given in Figure 4(c).

A recent overview of multigrid methods is required reading at this point. Yavneh (2006) introduced the method that is then explained for a number of relevant examples by Oosterlee and Gaspar-Lorenz (2006), Bergen et al (2006), and Bastian and Wieners (2006).

Obtain an approximate solution v_h at the finest grid, h

Calculate residual $r_h = F_h - K_h v_h$

Project residual by N levels to h-N

repeat: $r_{h-i} = R_{h-i}^{h-(i-1)} r_{h-i-1}$

Solve exactly

$\Delta v_{h-N} = K_{h-N} r_{h-N}$

Interpolate and improve

$r_{h-(i-1)} = R_{h-(i-1)}^{h-i} K_{h-i} \Delta v_{h-i}$

Improve $\Delta v_{h-(i-1)}$

$v_{h-(i-1)} = v_{h-(i-1)} + \Delta v_{h-(i-1)}$

Figure 5 A simple sawtooth multigrid algorithm for solving eqn [51].

7.05.4.2 Stokes' Flow: A Penalty Formulation

For 2-D problems, an efficient method to solve the incompressible Stokes' flow is a penalty formulation with a reduced and selective integration. This method has been widely used in 2-D thermal convection problems, for example, in ConMan (King et al., 1990). We now briefly discuss this penalty formulation, and detailed descriptions can be found in Hughes (2000).

The key feature in the penalty formulation is to allow for slight compressibility or $u_{k,k} \approx 0$. Here, it is helpful to make an analogy to isotropic elasticity. The constitutive equations for both compressible isotropic elasticity and incompressible isotropic elasticity are given by the following two equations:

$$\sigma_{ij} = -P\delta_{ij} + \eta\left(u_{i,j} + u_{j,i}\right) \quad [63]$$

$$u_{k,k} + P/\lambda = 0 \quad [64]$$

where λ is the Lamé constant that is finite for compressible media but infinite for incompressibility media (i.e., to satisfy $u_{k,k} = 0$ for finite P). To allow for slight compressibility, λ is taken finite but significantly larger than η, such that the error associated with the slight compressibility is negligibly small. Using words of 64 bit long (i.e., double precision), $\lambda/\eta \sim 10^7$ is effective. For finite λ, the constitutive equation becomes

$$\sigma_{ij} = \lambda u_{k,k}\delta_{ij} + \eta\left(u_{i,j} + u_{j,i}\right) \quad [65]$$

which replaces eqn [4].

An interesting consequence of this new constitutive equation is that the pressure is no longer needed in the momentum equation, and this simplifies the FE analysis. The weak form of the resulting Stokes' flow problem is

$$\int_\Omega w_{i,j} c_{ijkl} v_{k,l} \mathrm{d}\Omega = \int_\Omega w_i f_i \mathrm{d}\Omega + \sum_{i=1}^{n_{sd}} \int_{\Gamma_{h_i}} w_i h_i \mathrm{d}\Gamma - \int_\Omega w_{i,j} c_{ijkl} g_{k,l} \mathrm{d}\Omega \quad [66]$$

where

$$c_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \eta \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) \quad [67]$$

The FE implementation of eqn [66] is similar to that in Section 7.05.4.1. With the pressure excluded as a primary variable, the matrix equation is simply

$$[K]\{V\} = \{F\} \quad [68]$$

While the elemental force vector is defined the same as that in [44], the elemental stiffness needs some modification in comparison with that in [42].

$$k^e_{lm} = \vec{e}_i^{\,T} \left(\int_{\Omega^e} B_a{}^T D B_b \mathrm{d}\Omega + \int_{\Omega^e} B_a{}^T \overline{D} B_b \mathrm{d}\Omega \right) \vec{e}_j \quad [69]$$

where the first integral is the same as in [42] but the second integral is a new addition with

$$\overline{D} = \begin{bmatrix} \lambda & \lambda & 0 \\ \lambda & \lambda & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad [70]$$

The matrix eqn [68] only yields correct solution for velocities if a reduced and selective integration scheme is used to evaluate the elemental stiffness matrix (e.g., Hughes, 2000). Specifically, the numerical quadrature scheme for the second integral of eqn [69] needs to be one order lower than that used for the first integral. For example, if for a 2-D problem a 2×2 Gaussian quadrature scheme is used to evaluate the first integral, then a one-point Gaussian quadrature scheme is needed for the second integral. Hughes (2000) discussed the equivalence theorem for the mixed elements and the penalty formulation with the reduced and selective integration. Moresi et al. (1996) showed that these two formulations yield essentially identical results for Stokes' flow problems by comparing solutions from ConMan code employing a penalty formulation and Citcom that uses a mixed formulation.

Finally, we make two remarks about this penalty formulation.

First, although the pressure is not directly solved from the matrix equation, the pressure can be obtained through post-processing via $P = -\lambda u_{k,k}$ for each element. Such obtained pressure fields often display a checkerboard pattern. However, a pressure smoothing scheme (Hughes, 2000) seems to work well. The pressure field is important in many geophysical applications including computing dynamic topography and melt migration.

Second, with $\lambda/\eta \sim 10^7$, the stiffness matrix is not well conditioned and is not suited for any iterative solvers. A direct solver is required for this type of equations, as done in ConMan. This implies that this formulation may not be applicable to large 3-D problems because of the memory and computation requirements associated with direct solvers. Reducing λ/η improves the condition for the stiffness matrix; however, this is not recommended as it results in large errors associated with relaxing the incompressibility constraints.

7.05.4.3 The SUPG Formulation for the Energy Equation

The convective transport of any quantity at high Péclet number (the ratio of advective transport rate to diffusion rate) is challenging in any numerical approach in which the grid does not move with the material deformation. The transfer of quantities from grid points to integration points in order to calculate their updated values introduces a nonphysical additional diffusion term. Furthermore, the advection operator is difficult to stabilize, and many different schemes have been proposed to treat grid-based advection in both an accurate and stable fashion. Many of the successful approaches include some attempt to track the flow direction and recognize that the advection operator is not symmetrical in the upstream/downstream directions. This section introduces a SUPG formulation and a predictor–multicorrector explicit algorithm for time-dependent energy equation (i.e., eqn [3]). This method was developed by Hughes (2000) and Brooks (1981) some 20 years ago and remains an effective method in FE solutions of the equations with advection and diffusion such as our energy equation. FE mantle convection codes Citcom and ConMan both employ this method for solving the energy equation.

A weak form formulation for the energy eqn [3] and boundary conditions [7] is (Brooks, 1981)

$$\int_\Omega w\left(\dot{T} + u_i T_{,i}\right) \mathrm{d}\Omega + \int_\Omega w_{,i} (\kappa T_{,i}) \mathrm{d}\Omega + \sum_e \int_{\Omega_e} \overline{w} \left[\dot{T} + u_i T_{,i} - (\kappa T_{,i})_{,i} - \gamma \right] \mathrm{d}\Omega = \int_\Omega w \gamma \mathrm{d}\Omega + \int_{\Gamma_q} w q \mathrm{d}\Gamma - \int_\Omega w_{,i} \kappa g_{,i} \mathrm{d}\Omega \quad [71]$$

where w is the regular weighting functions and is zero on Γ_q, $\dot{T}$ is the time derivative of temperature, and $\overline{w}$ is the streamline upwind contribution to the weighting functions.

The FE implementation of [71] is similar to what was discussed for Stokes' flow in Section 7.05.4.1.2. While the weighting function w is similar to what was defined in [39] except it is now a scalar, the streamline upwind part $\overline{w}$ is defined through artificial diffusivity $\widetilde{\kappa}$ as

$$\overline{w} = \widetilde{\kappa} \hat{u}_j w_{,j} / |u| \quad [72]$$

where $|u|$ is the magnitude of flow velocity, $\hat{u}_j = u_j/|u|$ represents the directions of flow velocity, and $\widetilde{\kappa}$ is defined as

$$\widetilde{\kappa} = \left(\sum_{i=1}^{n_{sd}} \widetilde{\xi}_i u_i h_i \right) / 2 \quad [73]$$

$$\widetilde{\xi}_i = \begin{cases} 1 - 1/\alpha_i, & \alpha_i < -1 \\ 0, & -1 \le \alpha_i \le 1 \\ 1 - 1/\alpha_i, & \alpha_i > 1 \end{cases}, \quad \text{for } \alpha_i = \frac{u_i h_i}{2\kappa} \quad [74]$$

where u_i and h_i are flow velocity and element lengths in certain directions. It should be pointed out that eqns [72] and [74] are empirical and other forms are possible. Such defined streamline upward weighting function $\overline{w}$ can be thought as adding artificial diffusion to the actual diffusion term to lead to total diffusivity:

$$\kappa + \widetilde{\kappa} \hat{u}_i \hat{u}_j \quad [75]$$

$\overline{w}$ is discontinuous across elemental boundaries, different from w. This is why the integral in the third term of [71] is for each element. $\widetilde{w} = w + \overline{w}$ is also sometimes called the

Petrov–Galerkin weighting functions, which indicate that the shape function used to weight the integrals and the shape function for interpolation are distinct.

A reasonable assumption is that the weighted diffusion for an element in the third term of [71], $\overline{w}(\kappa T_{,i})_{,i}$, is negligibly small. Therefore, eqn [71] can be written as

$$\int_{\Omega} w_{,i}(\kappa T_{,i})\mathrm{d}\Omega + \sum_{e}\int_{\Omega_e}\widetilde{w}\left(\dot{T}+u_i T_{,i}-\gamma\right)\mathrm{d}\Omega = \int_{\Gamma_q} wq\mathrm{d}\Gamma - \int_{\Omega} w_{,i}\kappa g_{,i}\mathrm{d}\Omega \quad [76]$$

We now present relevant matrices at an element level. The $\dot{T}$ term in [71] implies that a mass matrix is needed and it is given as

$$m^e_{ab} = \int_{\Omega^e} N_a N_b \mathrm{d}\Omega \quad [77]$$

where $a,b = 1, \ldots, n_{en}$.

Elemental stiffness k^e is

$$k^e_{ab} = \int_{\Omega^e} {B_a}^T \kappa B_b \mathrm{d}\Omega \quad [78]$$

where for 2-D problems,

$${B_a}^T = \begin{pmatrix} N_{a,1} & N_{a,2} \end{pmatrix} \quad [79]$$

Elemental force vector f^e is given as

$$f^e_a = \int_{\Omega^e}\widetilde{N}_a\gamma\mathrm{d}\Omega + \int_{\Gamma^e_q} N_a q\mathrm{d}\Gamma - \sum_{b=1}^{n_{en}} k^e_{ab}g^e_b \quad [80]$$

where $\widetilde{N}_a$ is the Petrov–Galerkin shape function.

Elemental advection matrix c^e is given as

$$c^e_{ab} = \int_{\Omega^e}\widetilde{N}_a u_i N_{b,i}\mathrm{d}\Omega \quad [81]$$

The combined matrix equation may be written as

$$M\dot{\Phi} + (K+C)\Phi = F \quad [82]$$

where Φ is the unknown temperature and M, K, C, and F are the total mass, stiffness, advection matrices, and force vector assembled from all the elements.

Equation [82] can be solved using a predictor–corrector algorithm (Hughes, 2000) with some initial condition for temperature (e.g., eqn [5]). Suppose that temperature and its time derivative at time step n are given, Φ_n and $\dot{\Phi}_n$, the solutions at time step $n+1$ with time increment Δt can be obtained with the following algorithm:

(1) Predictor:

$$\Phi^0_{n+1} = \Phi_n + \Delta t(1-\alpha)\dot{\Phi}_n, \dot{\Phi}^0_{n+1} = 0 \quad [83]$$

for iteration step $i = 0$.

(2) Solving:

$$M^*\Delta\dot{\Phi}^i_{n+1} = \prod_e\left(f^e_{n+1} - m^e\dot{\Phi}^i_{n+1} - (k^e + c^e)\Phi^i_{n+1}\right) \quad [84]$$

(3) Corrector:

$$\Phi^{i+1}_{n+1} = \Phi^i_{n+1} + \Delta t\alpha\Delta\dot{\Phi}^i_{n+1}, \dot{\Phi}^{i+1}_{n+1} = \dot{\Phi}^i_{n+1} + \Delta\dot{\Phi}^i_{n+1} \quad [85]$$

(4) If needed, set iteration step $i = i+1$ and go back step 2.

We make four remarks for this algorithm. First, this method is second-order accurate if $\alpha = 0.5$ (Hughes, 2000). Second, typically two iterations are sufficient. Third, in [84], $\prod$ represents the operation of assembling elemental matrix into global matrix, and M^* is the lumped mass matrix that essentially makes this scheme an explicit scheme. Fourth, time increment Δt needs to satisfy Courant time-stepping constraints to make the scheme stable (Hughes, 2000).

7.05.5 Incorporating More Realistic Physics

In Section 7.05.2, we presented the governing equations for thermal convection in a homogeneous incompressible fluid with a Newtonian (linear) rheology and the Boussinesq approximation. However, the Earth's mantle is likely much more complicated with heterogeneous composition and non-Newtonian rheology (see Chapter 7.02). In addition, non-Boussinesq effects such as solid–solid phase transitions and mantle may play an important role in affecting the dynamics of the mantle. In this section, we will discuss the methods that help incorporate these more realistic physics in studies of mantle convection. We will focus on modeling thermochemical convection, solid-state phase transitions, non-Newtonian rheology, and mantle compressibility.

7.05.5.1 Thermochemical Convection

Thermal convection for a compositionally heterogeneous mantle has gained a lot of interest in recent years (Davaille, 1999; Kellogg et al., 1999; Lenardic and Kaula, 1993; Tackley, 1998a), with focus on the roles of mantle compositional anomalies and crustal structure in mantle dynamics. This is also called thermochemical convection. Different from purely thermal convection for which the fluid has the same composition, thermochemical convection involves fluids with different compositions. Here, we will present governing equations and numerical methods for solving these equations (see also Chapters 7.02 and 7.10).

7.05.5.1.1 Governing equations

Governing equations for thermochemical convection include a transport equation that describes the movement of compositions, in addition to the conservation laws of the mass, momentum, and energy (i.e., eqns [1]–[3]). Suppose that C describes the compositional field, the transport equation is

$$\frac{\partial C}{\partial t} + u_i C_{,i} = 0 \quad [86]$$

This transport equation is similar to the energy eqn [3] except that chemical diffusion and source terms are ignored, which is justified given that chemical diffusion is likely extremely small for the length scale and timecale that we consider in mantle convection. For a two-component system such as the crust–mantle system or depleted-primordial mantle system, C can be either 0 or 1, representing either component. If the fluids of different compositions have intrinsically different density, then the momentum eqn [2] needs to be modified to take into account of the compositional effects on the buoyancy

$$\sigma_{ij,j} + Ra(T - \beta C)\delta_{iz} = 0 \quad [87]$$

where β is the buoyancy number (Tackley and King, 2003; van Keken et al., 1997) and is defined as

$$\beta = \Delta\rho/(\rho\Delta T\alpha) \quad [88]$$

where $\Delta\rho$ is the density difference between the two compositions, ρ and ΔT are the reference density and temperature, and α is the reference coefficient of thermal expansion.

A special class of thermochemical convection problems examine how the mantle compositional heterogeneity is stirred by mantle convection (e.g., Christensen, 1989; Gurnis and Davies, 1986; Kellogg, 1992; van Keken and Zhong, 1999). For these studies on the mixing of the mantle, we may assume that the fluids with different composition have identical density with $\beta = 0$.

7.05.5.1.2 Solution approaches

Solving the conservation equations of the mass, momentum, and energy for thermochemical convection is identical to what was introduced in **Section 7.05.3** for purely thermal convection. The additional compositional buoyancy term in the momentum eqn [87] does not present any new difficulties numerically, provided that the composition C is given. The new challenge is to solve the transport eqn [86] effectively.

A number of techniques have been developed or adapted in solving the transport equation in thermochemical convection studies. They include a field method with a filter (e.g., Hansen and Yuen, 1988; Lenardic and Kaula, 1993), a marker chain method (Christensen and Yuen, 1984; van Keken et al., 1997; Zhong and Hager, 2003), and a particle method (e.g., Gerya and Yuen, 2003; McNamara and Zhong, 2004; Tackley, 1998a; Tackley and King, 2003; Weinberg and Schmeling, 1992). As reviewed by van Keken et al. (1997), while these techniques work to some extent, they also have their limitations, particularly in treating entrainment and numerical diffusion of composition C. We will briefly discuss each of these methods with more emphasis on the particle method.

In the particle method, the transport equation for C (i.e., eqn [86]) is not solved directly. Composition C at a given time is represented by a set of particles. This representation requires a mapping from the distribution of particles to compositional field C, which is often represented on a numerical mesh. With the mapping, to update C, all that is needed is to update the position of each particle to obtain an updated distribution of particles. This effectively solves the transport equation for C.

Two different particle methods have been used to map distribution of particles to C: absolute and ratio methods (Tackley and King, 2003). In the absolute method, particles are only used to represent one type of composition (e.g., for dense component or with $C = 1$). The population density of particles can be mapped to C. For example, C for an element/grid cell with volume Ω_e and particles N_e can be given as

$$C_e = AN_e/\Omega_e \quad [89]$$

where the constant A is the reciprocal of initial density of particles for composition $C = 1$ (i.e., total number of particles divided by the volume of composition $C = 1$). Clearly, the absence of particles in an element/grid cell represents $C = 0$. A physically unrealistic situation with $C > 1$ may arise due to statistical fluctuations in particle distribution or particle settling. Therefore, for this method to work effectively, a large number of particles are required (Tackley and King, 2003).

In the ratio method, two different types of particles are used to represent the compositional field C, type 1 for $C = 0$ and type 2 for $C = 1$. C for an element/grid cell that includes type 1 particles N_1 and type 2 particles N_2 is given as

$$C_e = N_2/(N_1 + N_2) \quad [90]$$

In the ratio method, C can never be greater than 1. Tackley and King (2003) found that the ratio method is particularly effective in modeling thermochemical convection in which the two components occupy similar amount of volumes.

We now discuss briefly procedures to update the positions of particles. One commonly used method is a high-order Runge–Kutta method (e.g., van Keken et al., 1997). Here, we present a predictor–corrector scheme for updating the particle positions (e.g., Zhong and Hager, 2003). Suppose that at time $t = t_0$, flow velocity is $\vec{u}_0$ and compositional field is C_0 that is defined by a set of particles with coordinates, $\vec{x}_0^i$, for particle i. The algorithm for solving composition at the next time step $t = t_0 + \mathrm{d}t = t_1$, C_1, can be summarized as follows:

(1) Using a forward Euler scheme, predict the new position for each particle i with $\vec{x}_{1p}^i = \vec{x}_0^i + \vec{u}_0 \mathrm{d}t$ and map the particles to compositional field C_{1p} at $t = t_1$.
(2) Using the predicted C_{1p}, solve Stokes' equation for new velocity $\vec{u}_{1p}$.
(3) Using a modified Euler scheme with second-order accuracy, compute the position for each particle i with $\vec{x}_1^i = \vec{x}_0^i + 0.5\left(\vec{u}_0 + \vec{u}_{1p}\right)\mathrm{d}t$ and compositional field C_1 at $t = t_1$.

The marker chain method is similar to the particle method in a number of ways. In the marker chain method, composition C is defined by an interfacial boundary that separates two different components. The interfacial boundary is a line for 2-D problems or a surface for 3-D. Using the flow velocity, one tracks the evolution of the interfacial boundary and hence composition C. Often, the interfacial boundary is represented by particles or markers. Therefore, updating the interfacial boundary is essentially the same as updating the particles in the particle method. Composition C on a numerical grid that is desired for solving the momentum and energy (eqns [87] and [3]) can be obtained by projection. As van Keken et al. (1997) indicated, the marker chain method is rather effective for compositional anomalies with relatively simple structure and geometry in 2-D.

The field method is probably the most straightforward. By setting diffusivity to be zero, we can employ the same solver for the energy equation (e.g., in **Section 7.05.4.3**) to solve the transport equation for C. However, this often introduces numerical artifacts including numerical oscillations and numerical diffusion. Lenardic and Kaula (1993) introduced a filter scheme that removes the numerical oscillations while conserving the total mass of compositional field.

Finally, the method of volume of fluid (Gopala and van Wachem, 2008; Hirt and Nichols, 1982) has been widely employed to investigate the dynamics of multiphase flow. The dynamics of a fluid phase injected into another fluid

medium of contrasting physical properties are a subject of wide interest in many physical disciplines, solid Earth geophysics being one of them. For example, the dynamics of fluid plumes, which are characterized by a balloon- or mushroom-shaped structure and evolve in pulses forming pinched-off structures in multiphase fluid systems (Webster and Longmire, 2001; Whitehead and Luther, 1975), can be best captured by this method. This method applies the flow equations based on volume averaging and tracks the interfaces of two fluid phases using a phase indicator function. It has been popularized by the commercial CFD code FLUENT (Fluent 6.0, 2001).

7.05.5.2 Solid-State Phase Transition

Solid-state phase transitions are important phenomena in the mantle. Major phase transitions include olivine–spinel transition at 410 km depth and spinel to perovskite and magnesium wüstite at 670 km depth, which are associated with significant changes in mantle density and seismic wave speeds. Recently, it was proposed that the D″ discontinuity near the core–mantle boundary is also caused by a phase transition from perovskite to postperovskite (Murakami et al., 2004). These phase transitions may affect the dynamics of mantle convection in two ways: (1) on the energetics due to latent heating associated with phase transitions and (2) on the buoyancy due to undulations at phase boundary caused by lateral variations in mantle temperature that affects the pressure at which phase transitions occur (Christensen and Yuen, 1985; Richter, 1973; Schubert et al., 1975; Tackley et al., 1993; Zhong and Gurnis, 1994). In this subsection, following Christensen and Yuen (1985) and Richter (1973), we present a method to model phase transitions.

The undulations of a phase boundary represent additional buoyancy force that affects the momentum equation. For phase transition k with density change $\Delta\rho_k$, the phase boundary undulations can be described by a dimensionless phase change function Γ_k that varies from 0 to 1 where regions with Γ_k of 0 and 1 represent the two phases separated by this phase change boundary. The momentum equation can be written as

$$\sigma_{ij,j} + \left(RaT - \sum_k Ra_k \Gamma_k\right)\delta_{iz} = 0 \qquad [91]$$

where phase change Rayleigh number Ra_k is

$$Ra_k = \frac{\Delta\rho_k g D^3}{\kappa\eta_0} \qquad [92]$$

The phase change function Γ_k is defined via 'excess pressure':

$$\pi_k = P - P_0 - \gamma_k T \qquad [93]$$

where γ_k and P_0 are the Clapeyron slope and phase change pressure at zero degree temperature for the kth phase transition. After normalizing pressure by $\rho_0 g D$ and Clapeyron slope by $\rho_0 g D/\Delta T$, the nondimensional 'excess pressure' can be written as

$$\pi_k = 1 - d_k - z - \gamma_k(T - T_k) \qquad [94]$$

where γ_k, d_k, and T_k are the nondimensional Clapeyron slope, reference phase transition depth, and reference phase transition temperature for the kth phase transition, respectively. The dimensionless phase change function is then given as

$$\Gamma_k = \frac{1}{2}\left(1 + \tanh\frac{\pi_k}{d}\right) \qquad [95]$$

where d is dimensionless phase transition width that measures the depth segment over which the phase change occurs. It should be pointed out that the effect of phase transition on buoyancy force can also be modeled with 'effective' coefficient of thermal expansion (Christensen and Yuen, 1985).

The latent heating effect along with viscous heating and adiabatic heating can be included in the nonlinear energy equation (also see eqn [3]) as (Christensen and Yuen, 1985)

$$\left[1 + \sum_k \gamma_k^2 \frac{Ra_k}{Ra}\frac{d\Gamma_k}{d\pi_k} D_i (T + T_s)\right]\left(\frac{\partial T}{\partial t} + \vec{v}\cdot\nabla T\right) + \left(1 + \sum_k \gamma_k \frac{Ra_k}{Ra}\frac{d\Gamma_k}{d\pi_k}\right)(T + T_s) D_i v_z = \nabla^2 T + \frac{D_i}{Ra}\tau_{ij}\frac{\partial v_i}{\partial x_j} + \gamma \qquad [96]$$

where T_s, v_z, and τ_{ij} are the surface temperature, vertical velocity, and deviatoric stress, respectively; k is phase change index; D_i is the dissipation number and is defined as

$$D_i = \frac{\alpha g D}{C_p} \qquad [97]$$

where α and C_p are the coefficient of thermal expansion and specific heat (see also **Chapter 7.02**). Christensen and Yuen (1985) called the effects of latent heating, viscous heating, and adiabatic heating as non-Boussinesq effects and termed this formulation as extended-Boussinesq formulation. They suggested that these effects are all of similar order, proportional to D_i, and should be considered simultaneously.

The modified momentum and energy (eqns [91] and [96]) can be solved with the same algorithms such as the Uzawa and SUPG for mantle convection problems with extended-Boussinesq approximations in 3-D (e.g., Kameyama and Yuen, 2006; Zhong, 2006).

7.05.5.3 Non-Newtonian Rheology

Laboratory studies suggest that the deformation of olivine, the main component in the upper mantle, follows a power-law rheology (e.g., Karato and Wu, 1993):

$$\dot{\varepsilon} = A\tau^n \qquad [98]$$

where $\dot{\varepsilon}$ is the strain rate, τ is the deviatoric stress, the pre-exponent constant A represents other effects such as grain size and water content, and the exponent n is ~3. The nonlinearity in the rheology arises from $n \neq 1$ (see also **Chapters 7.02** and **7.07**).

The effects of non-Newtonian rheology on mantle convection were first investigated by Parmentier et al. (1976), Cserepes (1982), and Christensen (1984). More recent efforts have been focused on how non-Newtonian rheology including viscoplastic rheology may lead to dynamic generation of plate tectonics (Bercovici, 1995; Foley and Becker, 2009; Jadamec and Billen, 2010; King and Hager, 1990; King et al., 2002; Moresi and Solomatov, 1998; Tackley, 1998b, 2000; Trompert and

Hansen, 1998; van Heck and Tackley, 2008; Zhong et al., 1998; see **Chapter 7.07**).

Solutions of nonlinear problems in general require an iterative approach. The power-law rheology may be written as an expression for effective viscosity:

$$\sigma_{ij} = -P\delta_{ij} + 2\eta_{\text{eff}}\dot{\varepsilon}_{ij} \quad [99]$$

$$\eta_{\text{eff}} = \tau/\dot{\varepsilon} = \frac{1}{A}\dot{\varepsilon}^{\frac{1-n}{n}} \quad [100]$$

where $\dot{\varepsilon}$ is given by the second invariant of strain-rate tensor

$$\dot{\varepsilon} = \left(\frac{1}{2}\dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij}\right)^{1/2} \quad [101]$$

It is clear that the effective viscosity depends on strain rate that in turn depends on flow velocity. Therefore, a general strategy for this problem is as follows: (1) Starting with some guessed effective viscosity, solve Stokes' flow problem for flow velocities; (2) update the effective viscosity with the newly determined strain rate, and solve Stokes' flow again; and (3) keep this iterative process until flow velocities are convergent.

Implementation of this iterative scheme is straightforward. The convergence for this iterative process depends on the exponent n. For regular power-law rheology with $n \sim 3$, convergence is usually not a problem. However, for large n (e.g., in case of viscoplastic rheology), the iteration may converge very slowly or may diverge. Often, some forms of damping may help improve convergence significantly (e.g., Jadamec et al., 2012; King and Hager, 1990).

7.05.5.4 Viscoelasticity

Although we primarily treat the mantle as a non-Newtonian viscous material, it may be required to consider viscoelastic behavior in cooler parts of the system such as the subducting lithosphere or on relatively short timescales such as mantle deformation following deglaciation for the last 10^4 years. While viscoelastic deformation is routinely considered for postglacial rebound studies (e.g., Peltier, 1976), its application to mantle convection is less common (Harder, 1991). To account for viscoelastic behavior, a formulation in terms of a perturbation to the viscous Stokes' equation is required. We consider the Maxwell viscoelastic behavior to illustrate the principles involved in the derivation and implementation.

In a Maxwell viscoelastic material, the constitutive law relates both shear *stress rate* and shear stress to strain rate. The assumption is that these deformation mechanisms work in parallel, that is,

$$\dot{\varepsilon}_{ij} = \frac{\dot{\tau}_{ij}}{2\mu} + \frac{\tau_{ij}}{2\eta} \quad [102]$$

where μ is the elastic shear modulus. We can approximate the stress rate with an FD expression such as

$$\dot{\tau}_{ij} = \frac{\tau_{ij} - \overset{\smile}{\tau}{}^{\Delta t^e}_{\text{ij}}}{\Delta t^e} \quad [103]$$

where Δt^e is the time step required to resolve the elastic effects correctly and $\overset{\smile}{\tau}{}^{\Delta t^e}_{ij}$ is the value of the stress for the fluid element at a time Δt^e earlier than the present time, t. Hence,

$$\dot{\varepsilon}_{ij} = \frac{\mu\Delta t^e + \eta}{2\mu\eta\Delta t^e}\tau_{ij} - \frac{\overset{\smile}{\tau}{}^{\Delta t^e}_{\text{ij}}}{2\mu\Delta t^e} \quad [104]$$

This can be written in terms of an effective viscosity η_{eff}

$$\tau_{ij} = 2\eta_{\text{eff}}\dot{\varepsilon}_{ij} + \eta_{\text{eff}}\frac{\overset{\smile}{\tau}{}^{\Delta t^e}_{\text{ij}}}{\mu\Delta t^e} \quad [105]$$

where $\eta_{\text{eff}} = \eta\mu\Delta t^e/(\eta + \mu\Delta t^e)$. Now, we substitute the deviatoric stress into the dimensional form of the momentum balance equation

$$\tau_{ij,j} - p_{,i} = g\rho\delta_{i3} \quad [106]$$

and note that the historical value, $\overset{\smile}{\tau}{}^{\Delta t^e}_{ij}$, is, in principle, a known value of the problem at time t and can be gathered onto the right-hand side:

$$\frac{\partial}{\partial x_j}\left[2\eta_{\text{eff}}\dot{\varepsilon}_{ij}\right] - p_{,i} = g\rho\delta_{i3} - \frac{\partial}{\partial x_j}\left[\eta_{\text{eff}}\frac{\overset{\smile}{\tau}{}^{\Delta t^e}_{\text{ij}}}{\mu\Delta t^e}\right] \quad [107]$$

This has the form of Stokes' equation but with a modified viscosity and stress history terms appearing as body forces. With these modifications, the solution strategies described earlier for creeping viscous flow are ideally suited for viscoelastic problems.

The deviatoric stress history, $\overset{\smile}{\tau}{}^{\Delta t^e}_{ij}$, is the value of the stress tensor at a previous time and, in a moving fluid, at a previous location and in a previous orientation (see **Figure 6**). There are two ways to compute the stress rate term, and one is to explicitly account for each of the contributing changes (i.e., time rate of change, translation, and rotation):

$$\dot{\tau}_{ij} = \frac{\partial\tau_{ij}}{\partial t} + u_k\frac{\partial\tau_{ij}}{\partial x_k} + \tau_{ik}W_{kj} - W_{ik}\tau_{kj} \quad [108]$$

where $W_{ij} = (\partial u_i/\partial x_j - \partial u_j/\partial x_i)/2$. This approach has the advantage that the terms are all evaluated at the computation point, but, as with scalar advection, specialized upwinding methods may be required to compute the terms in a stable and accurate manner. The second approach is to store an earlier value of the stress on a point and to track the translation and update the stored tensor value for rotation along the path of the fluid element. This latter approach is the basis for the various particle-based methods used in mantle/lithosphere dynamics (e.g., Gerya and Yuen, 2003; Moresi et al., 2003). It has the advantage of simplifying [108] to a simple ordinary differential equation (ODE) since all the transport terms are eliminated but introduces the need for tracking of fluid elements.

While [108], known as the Jaumann stress rate, is the minimal requirement for the stress rate to be objective (invariant to translation and rotation of the observer), other terms may also

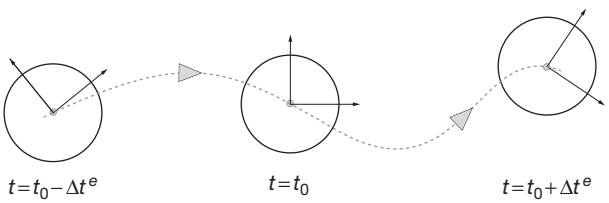


Figure 6 The rotation of the stress tensor during fluid flow.

be present to account, for example, for shearing associated with fluid flow. Muhlhaus and Regenauer-Lieb (2005) discussed various alternatives to the Jaumann rate in the context of mantle convection, and they pointed out that, at very high shear strain, the Jaumann rate produces nonphysical softening effects. They advocated ignoring the effects of rotation altogether in such cases. We point out, however, that there are many cases where rotations of stressed materials may be expected in the absence of substantial shear and would therefore caution our readers to understand and monitor carefully the situation in their own models (e.g., Bird et al., 1987).

We note that Δt^e is a parameter that needs to be chosen based on the physics of the problem and that also imposes an additional constraint on the minimum value of the time step. Maxwell materials are characterized by a *relaxation time*, which is defined by $t_r = \eta/\mu$, and the elastic discretization time, Δt^e, should be small enough to resolve the relaxation time. Elastic effects are significant when the relaxation time is larger than the characteristic time of the system. In the context of mantle convection, several possible definitions of the characteristic time are possible such as the inverse of the typical shear strain rate or a typical length scale divided by a typical velocity. In problems with a temperature- and stress-dependent viscosity, the relaxation time will vary within the system and so the appropriate choice of Δt^e is one that properly captures elastic behavior of *the region that influences the overall flow*. For example, in the hot mantle, the relaxation time is many orders of magnitude shorter than in the lithosphere, and elastic effects relax quickly compared to the timescale of flow. A choice of Δt^e that captures the relaxation time of elastic stresses in the cool lithosphere will produce a near-viscous response in the mantle as required since $\eta_{\text{eff}} \to \eta$ in the limit of zero relaxation time.

An effective viscosity formulation for viscoelastic–viscoplastic materials is also possible by introducing a third deformation mechanism in parallel with the viscous and elastic strain rates in [102] (Moresi et al., 2007a,b). It is also important to note that [103] is a simple first-order difference term that we use to illustrate the principle for introducing an elastic perturbation and is not expected to yield stable updates if elastic effects are likely to dominate the system. For a more complete description of viscoelastic rheology implemented for mantle dynamics problems, see Moresi et al. (2003, 2007a,b) and, for worked examples, see Gerya (2010). It is also worthwhile to point out that similar techniques to that discussed here have been used for incorporating the Maxwell viscoelastic rheology in mantle convection code CitcomS to study postglacial rebound, sea-level change, and time-varying gravity for the Earth's mantle with realistic 3-D viscoelastic structures (A et al., 2013; Paulson et al., 2005; Zhong et al., 2003).

7.05.5.5 Mantle Compressibility

Mantle compressibility has been suggested to have significant influence on plume dynamics for the Earth (e.g., Leng and Zhong, 2008a; Tan and Gurnis, 2005) and on mantle dynamics for recently discovered super-Earths (e.g, Liu and Zhong, 2013) that may have much larger mantle compressibility (e.g., Valencia et al., 2007). Mantle compressibility may also help compare mantle convection results with seismic and mineral physics models (e.g., Nakagawa et al., 2012). Therefore, there have been considerable developments of 2-D/3-D compressible mantle convection models with realistic rheology in recent years (e.g., King et al., 2010; Leng and Zhong, 2008b; Tan et al., 2011), although compressible mantle convection models had been formulated previously mostly for homogeneous mantle properties (Balachandar et al., 1995; Bercovici et al., 1992; Ita and King, 1998; Jarvis and McKenzie, 1980). For compressible convection under an anelastic-liquid approximation, the governing eqns [1]–[3] and rheological eqn [4] need to be modified (e.g., Leng and Zhong, 2008b; Schubert et al., 2001; see also **Chapter 7.02**). The nondimensional conservation equations of the mass and momentum become

$$(\rho_r u_i)_{,i} = 0 \qquad [109]$$

$$\sigma_{ij,j} + \left[\rho_r \alpha g Ra (T - T_r) - \frac{\alpha g}{c_P \Gamma} p\gamma\right]\delta_{iz} = 0 \qquad [110]$$

while the energy equation is similar to [96] with no phase changes. In [109] and [110], ρ_r represents a reference density profile determined from Adams–Williamson equation, Ra is defined by surface quantities, Γ is the Grüneisen parameter, γ is the mantle compressibility, and all other quantities are normalized by their corresponding surface values (e.g., Leng and Zhong, 2008b). Notice that the dynamic pressure's effect on buoyancy is considered to maintain the consistency of the energy balance (Leng and Zhong, 2008b). The rheological equation is

$$\sigma_{ij} = -p\delta_{ij} + \eta\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k}\right) \qquad [111]$$

Upon discretizing the governing eqns [109] and [110] with rheological eqn [111], we obtain matrix equations for Stokes' flow as (Leng and Zhong, 2008b)

$$KV + (G + W)P = F \qquad [112]$$

$$\left(G^T + C\right)V = 0 \qquad [113]$$

These two matrix equations are identical to [48] and [49] except for two terms: WP in [112] and CV in [113] that represent the dynamic pressure term in buoyancy and the product of vertical velocity and vertical gradient of density, respectively. Equations [112] and [113] can be solved using either biconjugate gradient method (e.g., Tan and Gurnis, 2007) or modified Uzawa method (Leng and Zhong, 2008b). In the latter method, the terms WP and CV are moved to the right-hand side and an iterative method is then applied to find a stable solution. The modified Uzawa method converges rapidly for a range of mantle compressibility (Leng and Zhong, 2008b) and has been implemented in CitcomS (Tan et al., 2011) and CitcomCU.

7.05.6 Advanced Numerical Techniques

Much of our discussion so far has focused on rather standard techniques used in computational geodynamics. While these techniques remain effective and relevant, we will discuss some more advanced techniques that have been developed in recent years and will see increasingly more applications in mantle convection studies. These techniques include more efficient

Stokes' flow solvers including treatment of nonlinearity and preconditioners and adaptive mesh refinement (AMR).

7.05.6.1 Adaptive Mesh Refinement

Multiscale physics is an important feature of mantle convection. The Earth's mantle convection is of very long wavelengths, for example, ~10 000 km for the Pacific Plate. However, mantle convection is also fundamentally controlled by thin thermal boundary layers that lead to thin upwelling plumes (~100 km) and downwelling slabs due to high Rayleigh number in mantle convection. Plate boundary processes and entrainment in thermochemical convection also occur over possibly even smaller length scale. Furthermore, material properties are also affected by near microscopic properties such as grain size, which has not been well incorporated in mantle convection studies (see also **Chapter 7.07**). Most existing codes in mantle convection (e.g., CitcomS and STAG3D) work for largely uniform grids, and the largest runs to date have used several hundred million unknowns (see Jadamec and Billen, 2010). Recent codes such as ASPECT (Kronbichler et al., 2012) and Rhea (Burstedde et al., 2008, 2013) use dynamic AMR (Davies et al., 2007, 2011; Leng and Zhong, 2011; May et al., 2013). These codes, fully parallelized for massively parallel computers (1000 s to 100 000 cores), may have numerical resolution as high as a few kilometers locally where needed even in 3-D global models and therefore offer great opportunities to solve some of the computationally challenging and demanding geodynamics questions (e.g., Stadler et al., 2010). It is most appropriate for problems with very localized heterogeneity, rather than small features across the entire domain.

AMR introduces significant topological and mathematical complexity to the numerical analysis and codes. It requires efficient algorithms for mesh refinement, coarse, readaptation, and repartition, and it also demands efficient parallel partitioning methods to distribute computational work equally among processors. To deal with these challenges, Rhea uses a technique called a forest of octrees in which the computational domain is divided into subdomains each with an octree structure (Burstedde et al., 2008, 2013). An octree is a logical structure in which each node either is a leaf or has child nodes. The structure is recursive and hierarchical and can be implemented efficiently for FE analyses and parallel computing, as in Rhea. ASPECT is built on numerical software package deal.II with higher-order spatial and temporal discretization schemes, and it also uses a forest of octrees for parallel computing. Both Rhea and ASPECT use some forms of explicit algorithms for time stepping in solving the energy equation. A challenge for time-dependent mantle convection problems with kilometer resolution is the relatively small time step required for numerical stability.

7.05.6.2 Nonlinear Solver

With a stress-dependent non-Newtonian rheology or plasticity, Stokes' flow problem becomes nonlinear (see **Section 7.05.5.3**). Thus, an extra outer iteration is necessary to solve the system. The simplest approach modifies the original Stokes' solver by freezing the viscosity at its current value for the given velocity, solving the resulting Stokes' system to update the velocity, and iterating until convergence. This iteration is commonly known as the *Picard method* and has been commonly used in geodynamics (e.g., for Citcom (e.g., Zhong et al., 1998) and ConMan codes (King and Hager, 1990)). The Picard method is limited to a linear convergence rate, which may be quite slow, whereas Newton's method can achieve quadratic convergence. Newton's method (Dembo et al., 1982) requires a solution with the Jacobian J of the nonlinear operator to calculate each update δu:

$$J(u)\delta u = -F(u) \qquad [114]$$

A naive implementation of Newton's method may fail to converge for starting guesses far from the solution. It is commonly augmented with a line search parameter λ:

$$u_{\text{new}} = u + \lambda\delta u \qquad [115]$$

where λ is chosen to minimize the residual $F(u_{\text{new}})$ along the update direction. In addition, a few steps of the Picard method can be used to produce a better starting guess for Newton's method. It is possible to generalize this strategy by allowing more sophisticated combinations of nonlinear solvers, or *nonlinear preconditioning* (Brune et al., 2013). In particular, we can use nonlinear accelerators such as nonlinear GMRES (Washio and Oosterlee, 2000) to greatly expand the region of convergence of Newton's method.

The construction of a full Jacobian is costly in terms of both memory and time. Quasi-Newton methods (Dennis and More, 1977) can be used to reduce memory usage since they approximate the action of the Jacobian using a few vectors, improving the approximation at each iterate. This strategy can also be used to reduce construction time by allowing recalculation of the Jacobian only after several iterations of the Newton algorithm without slowing the convergence rate (Brown and Brune, 2013).

Following from the discussion in **Section 7.05.4.1.3**, we may invert the relationship between block decomposition and multigrid for the nonlinear problem. The full system can be solved using the full approximation scheme (FAS; Brandt, 1977) or nonlinear multigrid, and for the smoother, we may use a Newton method with a block preconditioner.

7.05.6.3 The Preconditioned Uzawa Algorithm

The convergence rates of Krylov methods in the Uzawa algorithm for Stokes' flow problem (**Section 7.05.4.1.3**) are related to the condition number of the matrix $\hat{K}$ (see explanation in the succeeding text). Systems with a large condition number that may arise from spatial variations in viscosity or uneven meshes are considered ill-conditioned, and, in this case, the rate of convergence of the iteration can be very slow. It is possible, however, to improve this situation dramatically by introducing appropriate 'preconditioning' of the system. We attempt to transform the problem to one in which the condition number is reduced using a transformation, which is simple and cheap to apply.

For our matrix equation [50], that is, $\hat{K}P = \hat{F}$, let us consider a related problem $\hat{K}'P' = \hat{F}'$ where

$$\hat{K}'P' = \left(C^{-1}\hat{K}C^{-1}\right)\cdot(CP) = \left(C^{-1}\hat{F}\right) = \hat{F}' \qquad [116]$$

where C is a symmetrical positive definite matrix that has the property that the condition number of $\hat{K}'$ is much improved over that of $\hat{K}$. Once again, we use the conjugate gradient algorithm as our example Krylov solver. When [116] is substituted into the conjugate gradient algorithm, some straightforward rearrangement (following Golub and van Loan, 1989) allows us to deal only with the action of the matrix $M = C^2$ and the untransformed matrix/vectors. M is known as the preconditioner and the resulting preconditioned form of conjugate gradient algorithm is shown in **Figure 3(b)**.

For this process to be useful, the solution time for $Mz_k = r_k$ needs to be small relative to the time saved by reducing the number of iterations. In the limit that M approaches $\hat{K}$, only a single iteration is required. Hence, in an informal sense, a good preconditioner is one that approximates $\hat{K}$ but is easier to solve. To be more precise, a good choice of M should have a comparable eigenvalue spectrum to $\hat{K}$. The choice of preconditioner used for Citcom codes is

$$M_{\text{CITCOM}} = \text{diag}\left[G^T \text{diag}[K]^{-1} G\right] \qquad [117]$$

which can be constructed and inverted trivially. The entries in diag[K] are the average viscosity in the elements associated with a given degree of freedom. In Stokes' flow problem with variable viscosity, this trivial approximation can greatly improve the convergence of the iteration (Moresi and Solomatov, 1995). This is an example of a preconditioner that has an obvious physical meaning but is also constructed algebraically from the original matrix. For a more detailed explanation of preconditioning of this particular system, including more sophisticated constructions of M, we refer the reader to May and Moresi (2008).

When we say that a system of equations is ill-conditioned, we mean precisely that the *condition number* is large. The condition number of a linear system A

$$\kappa(A) = \|A\|\left\|A^{-1}\right\| = \lambda_1/\lambda_N \qquad [118]$$

is the measure of how close the system is to unsolvable and also the ratio of the largest to the smallest eigenvalue. This in turn gives a good idea of its difficulty to solve iteratively. Although very expensive to calculate exactly, the condition number can be efficiently approximated using the same Krylov iterations that we use for solving it. The projection of A into the Krylov subspace is a small matrix, H, for which we can easily calculate $\kappa(H) \approx \kappa(A)$. Thus, a few iterations of the Krylov method provide us with a monitor of the system difficulty. This same methodology is also used to provide spectral bounds for polynomial smoothers, such as the Chebyshev iteration, in multigrid smoothers discussed in **Section 7.05.4.1.4**.

7.05.6.4 Libraries

Simulation codes now have a solid software foundation of high-quality numerical libraries on which to build. Incorporating libraries into development can improve not only the code performance but also the portability, maintainability, and extensibility of the application itself.

The most well-developed scientific libraries are those for linear algebra, both dense (e.g., Elemental, Poulson et al., 2013) and sparse (e.g., PETSc, Balay et al., 2013a,b; Katz et al., 2007) including recent additions for massively parallel architectures (e.g., ViennaCL and VexCL, Demidov et al., 2012; Rupp et al., 2010), which benefit from a clear mathematical foundation, widely accepted interface, and long history of careful optimization. Solvers for both linear and nonlinear algebraic equations are also highly developed (e.g., PETSc, Brown et al., 2012, Trilinos, Heroux et al., 2013) including scalable solvers such as multigrid. Specific libraries for multigrid solvers are also available, such as ML (Hu et al., 2013), Hypre (Falgout et al., 2013), and OpenCurrent (Cohen, 2013). However, there are now libraries available with interfaces for high-level objects, such as FE weak forms (e.g., FEniCS, Logg et al., 2011; Wells et al., 2012), deal.II (Bangerth et al., 2007, 2013)), and more general computation fluid dynamics (e.g., OpenFOAM, OpenFOAM, 2013, and Fluidity, ICL, 2013). Mantle dynamics questions have also been answered using libraries with wider capabilities. Underworld (Capitano et al., 2010; Moresi et al., 2007a) and Escript (Gross et al., 2007), which can simulate full viscoplastic rheologies, have been used to investigate, for example, the dynamics of oceanic trench migration.

Recently, fully developed applications for mantle dynamics questions have emerged that are built using the libraries mentioned earlier as a base. LaMEM (Keller et al., 2013), following the design of MILAMIN (Dabrowski et al., 2008), incorporates the PETSc library for parallel data management and algebraic solvers. TerraFERMA, capable of solving the full magma dynamics problems, uses the FEniCS library for FE modeling and PETSc in much the same way as LaMIN. We anticipate that this style of software construction will become more prevalent as code complexity increases and library development techniques become better established.

7.05.6.5 Computer Architecture

A revolution in computer architecture has replaced the frequency scaling, relied upon for decades to keep up with Moore's law predictions, with a drive toward parallelism. Now, continued Moore's law scaling and reasonable power consumption are only possible by packing more cores on a chip. Currently, a majority of the ten fastest machines in the world rely on massively parallel accelerators (Jeffers and Reinders, 2013; Meuer et al., 2001; Yuen et al., 2013).

Modern architectures are evolving toward higher concurrency (i.e., more computations happening simultaneously in groups of threads) and much smaller local memory. In addition, the memory hierarchy is much more pronounced, with global memory access several hundred times more expensive than local accesses. Concurrency is necessary to increase processing power because it is prohibitively expensive, in terms of power, to increase the speed of an individual processor. Also, memory transfers consume the lion's share of power in a processor, and transferring more memory per request is the only way to increase memory bandwidth. Therefore, simulation codes will need to cope with increasing demands for parallelism for concurrent execution and vectorization to use the large blocks of requested memory efficiently.

This should not necessitate the wholesale change of simulation codes, but rather the targeted use of libraries that are optimized for these architectures. For example, the PETSc linear algebra interface can make use of the ViennaCL library for massively concurrent execution on accelerators such as GPUs and Intel MIC. Higher-level operations, such as Krylov methods, are built from these components. However, some parts of user code such as residual evaluation (Knepley and Terrel, 2013; Knepley et al., 2013) will likely have to be reengineered so that element traversals are suitable for these new machines (Yuen et al., 2013).

7.05.6.6 Boundary Element Methods

Constant coefficient elliptical equations admit a Green function representation, so that the equation to be solved is a Fredholm integral equation. These equations can be discretized using a boundary element formalism. This strategy can be attractive due to the simplification in workflow for the simulation, since only the boundary need be meshed. In the case of moving boundaries, such as descending slabs, this can be a vast simplification. Moreover, the integral equations arising from this discretization very often have condition number independent of the grid size so that simple solvers can be used. Boundary element solvers have been used to model subduction (Morra et al., 2010), overriding plates (Dinther et al., 2010), and oceanic arcs (Morra et al., 2006).

Unlike the FD or FE discretizations of a differential operator, the boundary element discretization of the integral operator does not produce a sparse matrix. Since the integral operator is typically well conditioned, it is enough to use a simple Krylov iteration for the solution, and thus, only the action of the operator is necessary. The fast multipole method (Greengard and Rokhlin, 1987) can be used to calculate this action in time linear in the number of boundary elements, rather than the quadratic time necessary if the dense operator is fully formed, as well as other related methods (Phillips and White, 1997). These approaches have been encapsulated in libraries (Altman et al., 2006; Cruz et al., 2010; Ying et al., 2003).

7.05.7 Concluding Remarks and Future Prospect

In this chapter, we have discussed four basic numerical methods for solving mantle convection problems: FE, FD, FV, and spectral methods. We have focused our efforts on FE method, mainly because of its growing popularity in mantle convection studies over the past decade, partially prompted by the easily accessed FE codes from ConMan to Citcom (www.geodynamics.org). To this end, the discussions on FE method should help readers understand the inner working of these two FEs codes. However, our discussions on FD, FV, and spectral methods are rather brief and are meant to give readers a solid basis for understanding the rudiments of these methods and the references with which to delve deeper into the subjects. These three methods have all been widely used in studies of mantle convection and will most likely remain so for years to come. It is our view that each of these methods has its advantages and disadvantages and readers need to find the one that is most suited to their research. This chapter is by no means exhaustive or extremely advanced in character. We did not cover many potentially interesting and powerful numerical techniques, such as spectral elements (e.g., Komatitsch and Tromp, 1999), radial basis functions method (Flyer and Wright, 2009; Wright et al., 2010), wavelets (Daubeschies et al., 1985), level-set method (Osher and Fedkiw, 2003), mesh-free smoothed particle hydrodynamics method (Gingold and Monaghan, 1977; Hoover and Hoover, 2001), and adaptive grid techniques (Berger and Oliger, 1984; Bruegmann and Tichy, 2004), and interested readers can read these references to learn more.

Rapid advancement in computing power has made 3-D modeling of mantle convection practical, although 2-D modeling still plays an essential role. While a variety of solvers with either iterative or direct solution method are available to 2-D models, 3-D modeling requires iterative solution techniques due to both computer memory and computational requirements. A powerful iterative solution approach is the multigrid methods that can be used in either FE or FV method. Such methods are already implemented in a number of mantle convection codes including STAG3D (Tackley, 1994), Citcom/CitcomS (Moresi and Solomatov, 1995; Zhong et al., 2000, 2008), Terra (Baumgardner, 1985), and others (Kameyama et al., 2008; Tackley, 2008). We spent some effort in discussing the multigrid method, and more on this topic can be found in Brandt (1977), Trottenberg et al. (2001), and Yavneh (2006). The multigrid idea is powerful in that one can generalize this to structures other than grids, for instance, multiscale or multilevel techniques (Trottenberg et al., 2001).

3-D modeling should almost certainly make a good use of parallel computing with widely available parallel computers from PC clusters to super–parallel computers with well over 3,000,000 cores, such as the Milky Way-2 (Liao et al., 2014). Parallel computing technology poses certain limitations on numerical techniques as well. For example, spectral methods, while having some important advantages over other grid-based numerical methods, are more difficult to implement efficiently for parallel computers and to incorporate geometric complexity. This may severely limit its use to tackle next-generation computing problems. Fortunately, many other numerical methods including FE, FD, and FV methods are very efficient for parallel computing. Many of the codes mentioned earlier for mantle convection studies including STAG3D, Citcom/CitcomS, Terra, ASPECT, Rhea, and those in Harder and Hansen (2005) and Kageyama and Sato (2004) are fully parallelized and can be used on different parallel computers. These codes can be scaled up to possibly 10 000s of cores with a great potential yet to be explored in helping understand high-resolution, high Rayleigh number mantle convection. Beyond 10 000 cores, one would likely need to employ specialized libraries, such as PETSc and p4est, which have been shown to scale to the largest machines.

However, there remain many challenges in numerical modeling of mantle convection, both in developing more robust numerical algorithms and in analyzing model results. At least two areas need better numerical algorithms:

(1) Thermochemical convection becomes increasingly important in answering a variety of geodynamic questions.

We discussed cursorily a Lagrangian technique of advection of passive tracers in solving thermochemical convection. However, it is clear that most existing techniques do not work well for entrainment in thermochemical convection, as demonstrated by van Keken et al. (1997) and Tackley and King (2003). The increasing computing power will help solve this problem by providing significantly high resolution, but better algorithms are certainly needed.

(2) The lithosphere is characterized by highly nonlinear rheology including complex shear-localizing feedback mechanisms and history-dependent rheology and plastic deformation (Bercovici, 2003; Gurnis et al., 2000; see **Chapter 7.07**). Convergence deteriorates rapidly when nonlinearity increases. More robust algorithms are needed for solving mantle convection with highly nonlinear rheology.

Efficient postprocessing and analyses of modeling results are also increasingly becoming an important issue. Mantle convection, along with many other disciplines in the geosciences, now faces an exponential increase in the amount of numerical data generated in large-scale high-resolution 3-D convection. It is currently commonplace to have a few terabytes of data for a single project, and this poses significant challenges to conventional ways of data analyses, postprocessing, and visualization. Visualization is already a severe problem in the era of petascale computing. Considering our future quest to the exascale computing, this problem would be greatly exacerbated in the next decade. It is essential to employ modern visualization tools to confront this challenge. Erlebacher et al. (2001) discussed these issues and their solutions, and similar arguments about these problems and their solutions can be found in a report on high-performance computing in geophysics by Cohen (2005). Several potentially useful methodologies such as 2-D/3-D feature extraction, segmentation methods, and flow topology (see Hansen and Johnson, 2005; Kreylos et al., 2006) can help geophysicists understand better the physical structure of time-dependent convection with coherent structures, such as plumes or detached slabs in spherical geometry. 3-D visualization packages, for example, Amira or ParaView, can help to alleviate the burden of the researcher in unraveling the model output. Going further into detailed examination of 3-D mantle convection, one would need large-scale display devices such as the power walls with more than 30 million pixels or CAVE-like environments. This particular visualization method is described under current operating conditions in the Earth Simulator Center by Ohno et al. (2006) and other sites around the world. Remote visualization of the data under the auspices of Web services using the client–server paradigm may be a panacea for large-scale collaborative projects (see Erlebacher et al., 2006).

Finally, it is vitally important to develop and maintain efficient and robust benchmarks, as in any fields of computational sciences. Benchmark efforts have been made in the past by various groups for different mantle convection problems (e.g., Blankenbach et al., 1989; King et al., 2010; Moresi and Solomatov, 1995; Stemmer et al., 2006; Tackley, 1994; van Keken et al., 1997; Zhong et al., 2000, 2008). As numerical methods and computer codes become more sophisticated and our community moves into the petascale computing era, benchmark efforts become even more important to assure the efficiency, scalability, and accuracy of numerical models of mantle convection.

References

A GR, Wahr J, and Zhong SJ (2013) Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: An application to glacial isostatic adjustment in Antarctica and Canada. *Geophysical Journal International* 192: 557–572.

Albers M (2000) A local mesh refinement multigrid method for 3-D convection problems with strongly variable viscosity. *Journal of Computational Physics* 160: 126–150.

Altman MD, Bardhan JP, Tidor B, and White JK (2006) FFTSVD: A fast multiscale boundary-element solver suitable for Bio-MEMS and biomolecule simulation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 25(2): 274–284.

Atanga J and Silvester D (1992) Iterative methods for stabilized mixed velocity-pressure finite elements. *International Journal for Numerical Methods in Fluids* 14: 71–81.

Auth C and Harder H (1999) Multigrid solution of convection problems with strongly variable viscosity. *Geophysical Journal International* 137: 793–804.

Balachandar S and Yuen DA (1994) Three-dimensional fully spectral numerical method for mantle convection with depth-dependent properties. *Journal of Computational Physics* 132: 62–74.

Balachandar S, Yuen DA, Reuteler DM, and Lauer G (1995) Viscous dissipation in three dimensional convection with temperature-dependent viscosity. *Science* 267: 1150–1153.

Balay S, Brown J, Buschelman K, et al. (2013a) PETSc Web page, http://www.mcs.anl.gov/petsc.

Balay S, Brown J, Buschelman K, et al. (2013b) PETSc users manual, Technical Report ANL-95/11 – Revision 3.4. Argonne National Laboratory.

Bangerth W, Hartmann R, and Kanschat G (2007) deal.II—A general-purpose object-oriented finite element library. *ACM Transactions on Mathematical Software* 33(4).

Bangerth W, et al. (2013) deal.ii Web Page, http://www.dealii.org.

Bastian P and Wieners C (2006) Multigrid methods on adaptively refined grids. *Computing in Science and Engineering* 8(6): 44–55.

Bathe K-J (1996) *Finite Element Procedures*. Englewood Cliffs, NY: Prentice Hall, 1037 pp.

Baumgardner JR (1985) Three dimensional treatment of convection flow in the Earth's mantle. *Journal of Statistical Physics* 39(5/6): 501–511.

Bercovici D (1995) A source-sink model of the generation of plate tectonics from non-Newtonian mantle flow. *Journal of Geophysical Research* 100: 2013–2030.

Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121.

Bercovici D, Schubert G, and Glatzmaier GA (1992) Three-dimensional convection of an infinite-Prandtl number compressible fluid in a basally heated spherical shell. *Journal of Fluid Mechanics* 239: 683–719.

Bergen B, Gradl T, Huelsemann F, and Ruede U (2006) A massively parallel multigrid method for finite elements. *Computing in Science and Engineering* 8(6): 56–62.

Berger M and Oliger J (1984) Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of Computational Physics* 53: 484–512.

Bird RB, Armstrong RC, and Hassager O (1987) Dynamics of polymeric fluids. *Fluid Mechanics*, vol. 1 New York: Wiley.

Blankenbach B, Busse F, Christensen U, et al. (1989) A benchmark comparison of mantle convection codes. *Geophysical Journal International* 98: 23–38.

Brandt A (1977) Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation* 31: 333–390.

Briggs WL, Henson VE, and McCormick SF (2000) *A Multigrid Tutorial*, 2nd edn. SIAM Press.

Brooks AN (1981) *A Petrov–Galerkin Finite Element Formulation for Convection Dominated Flows*. PhD Thesis, Pasadena, CA: California Institute of Technology.

Brown J and Brune P (2013) Low-rank quasi-Newton updates for robust Jacobian lagging in Newton-type methods. In: *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering*, pp. 2554–2565.

Brown J, Knepley MG, May DA, McInnes LC, and Smith BF (2012) Composable linear solvers for multiphysics. In: *Proceedings of the 11th International Symposium on Parallel and Distributed Computing (ISPDC 2012)*, pp. 55–62. IEEE Computer Society.

Bruegmann B and Tichy W (2004) Numerical solutions of orbiting black holes. *Physical Review Letters* 92: 211101.
Brune P, Knepley MG, Smith B, and Tu X (2013) *Composing Scalable Nonlinear Algebraic Solvers*. Preprint ANL/MCS-P2010-0112, Argonne National Laboratory.
Bunge H-P, Richards MA, and Baumgardner JR (1997) A sensitive study of 3-dimensional spherical mantle convection at 108 Rayleigh number: Effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *Journal of Geophysical Research* 102: 11991–12007.
Bunge HP, Richards MA, Lithgow-Bertelloni C, Baumgardner JR, Grand SP, and Romanowicz BA (1998) Time scales and heterogeneous structure in geodynamic earth models. *Science* 280: 91–95.
Burstedde C, Stadler G, Alisic L, Wilcox LC, Tan E, Gurnis M, et al. (2013) Large-scale adaptive mantle convection simulations. *Geophysical Journal International* 192: 889–906.
Burstedde C, Ghattas O, Gurnis M, et al. (2008) Scalable parallel mantle convection simulation on petascale computers. In: *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, Austin, TX*. http://dx.doi.org/10.1109/SC.2008.5214248.
Busse FH (1989) Fundamentals of thermal convection. In: Peltier WR (ed.) *Mantle Convection, Plate Tectonics and Global Dynamics*, pp. 23–109. Gordon and Breach.
Capitano FA, Stegman DR, Moresi LN, and Sharples W (2010) Upper plate controls on deep subduction, trench migration, and deformations at convergent margins. *Tectonophysics* 483(1–2): 80–92.
Chandrasekhar S (1961) Hydrodynamic and hydromagnetic stability. *International Series of Monographs on Physics* Oxford: Clarendon Press.
Choblet G, Cadek O, Couturier F, and Dumoulin C (2007) Oedipus: A new tool to study the dynamics of planetary interiors. *Geophysical Journal International* 170: 9–30. http://dx.doi.org/10.1111/j.1365-246X.2007.03419.x.
Christensen UR (1984) Convection with pressure-and temperature-dependent non-Newtonian rheology. *Geophysical Journal of the Royal Astronomical Society* 77: 343–384.
Christensen U (1989) Mixing by time-dependent convection. *Earth and Planetary Science Letters* 95: 382–394.
Christensen U and Harder H (1991) Three-dimensional convection with variable viscosity. *Geophysical Journal International* 104: 213–226.
Christensen UR and Yuen DA (1984) The interaction of a subducting lithosphere with a chemical or phase boundary. *Journal of Geophysical Research* 89(B6): 4389–4402.
Christensen UR and Yuen DA (1985) Layered convection induced by phase changes. *Journal of Geophysical Research* 90(10): 10291–10300.
Cohen RE (ed.) (2005) *High-Performance Computing Requirements for the Computational Solid-Earth Sciences*, 101 pp, http://www.geoprose.com/computational_SES.html.
Cohen J (2013) Open Current Web page, https://code.google.com/p/opencurrent.
Cruz F, Knepley MG, and Barba LA (2010) PetFMM—A dynamically load-balancing parallel fast multipole library. *International Journal for Numerical Methods in Engineering* 85(4): 403–428.
Cserepes L (1982) Numerical studies of non-Newtonian mantle convection. *Physics of the Earth and Planetary Interiors* 30: 49–61.
Cserepes L and Rabinowicz M (1985) Gravity and convection in a two-layer mantle. *Earth and Planetary Science Letters* 76: 193–207.
Cserepes L, Rabinowicz M, and Rosemberg-Borot C (1988) Three-dimensional infinite Prandtl number convection in one and two layers with implication for the Earth's gravity field. *Journal of Geophysical Research* 93: 12009–12025.
Dabrowski M, Krotkiewski M, and Schmid DW (2008) MILAMIN: Matlab-based finite element method solver for large problems. *Geochemistry, Geophysics, Geosystems* 9, Q04030.
Daubeschies I, Grossmann A, and Meyer Y (1985) Painless nonorthogonal expansions. *Journal of Mathematical Physics* 27: 1271–1283.
Davaille A (1999) Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* 402: 756–760.
Davies GF (1995) Penetration of plates and plumes through the mantle transition zone. *Earth and Planetary Science Letters* 133: 507–516.
Davies DR, Davies JH, Hassan O, Morgan K, and Nithiarasu P (2007) Investigations into the applicability of adaptive finite element methods to two-dimensional infinite Prandtl number thermal and thermochemical convection. *Geochemistry, Geophysics, Geosystems* 8: Q05010. http://dx.doi.org/10.1029/2006GC001470.
Davies DR, Wilson CR, and Kramer SC (2011) Fluidity: A fully unstructured anisotropic adaptive mesh computational modeling framework for geodynamics. *Geochemistry, Geophysics, Geosystems* 12(6): Q06001. http://dx.doi.org/10.1029/2011GC003551.
Dembo R, Eisenstat S, and Steihaug T (1982) Inexact Newton methods. *SIAM Journal on Numerical Analysis* 19(2): 400–408.
Demidov D, Ahnert K, Rupp K, and Gottschling P (2012) *Programming CUDA and OpenCL: A Case Study Using Modern C++ Libraries*, ArXiv e-prints.
Dennis JE Jr. and More J (1977) Quasi-Newton methods, motivation and theory. *SIAM Review* 19(1): 46–89.
Desai CS and Abel JF (1972) Introduction to the finite element method. In: *A Numerical Method for Engineering Analysis*. New York: Academic Press.
Dinther Y, Morra G, Funiciello F, and Faccenna C (2010) Role of the overriding plate in the subduction process: Insights from numerical models. *Tectonophysics* 484: 74–86.
Erlebacher C, Yuen DA, and Dubuffet F (2001) Current trends and demands in visualization in the geosciences. *Visual Geosciences* 6. http://dx.doi.org/10.1007/s10069-001-1019-y, 59 pp.
Erlebacher G, Yuen DA, Lu Z, Bollig EF, Pierce M, and Pallickara S (2006) A grid framework for visualization services in the Earth Sciences. *Pure and Applied Geophysics* 163: 2467–2483.
Eshof J, Sleijpen G, and Gijzen M (2005) Relaxation strategies for nested Krylov methods. *Journal of Computational and Applied Mathematics* 177(2): 347–365.
Falgout R, et al. (2013) Hypre Web page, http://computation.llnl.gov/casc/linear_solvers/sls_hypre.html.
Fluent 6.0 Users Guide Documentation (2001) Fluent Inc.
Flyer N and Wright GB (2009) A simple radial basis function method for the shallow water equations on a sphere. *Proceedings of the Royal Society A* 465: 1949–1976.
Foley B and Becker TW (2009) Generation of plate-like behavior and mantle heterogeneity from a spherical, visco-plastic convection model. *Geochemistry, Geophysics, Geosystems* 10: Q08001. http://dx.doi.org/10.1029/2009GC002378.
Fornberg B (1990) High-order finite differences and the pseudospectral method on staggered grids. *SIAM Journal on Numerical Analysis* 27: 904–918.
Fornberg B (1995) *A Practical Guide to Pseudospectral Methods*. Cambridge University Press.
Gable CW, O'Connell RJ, and Travis BJ (1991) Convection in three dimensions with surface plates: Generation of toroidal flow. *Journal of Geophysical Research* 96: 8391–8405.
Gerya T (2010) *Introduction to Numerical Geodynamic Modeling*. Cambridge: Cambridge University Press.
Gerya TV and Yuen DA (2003) Characteristics-based marker method with conservative finite-differences schemes for modeling geological flows with strongly variable transport properties. *Physics of the Earth and Planetary Interiors* 140: 295–320.
Gerya TV and Yuen DA (2007) Robust characteristics method for modeling multiphase visco-elastoplastic thermo-mechanical problems. *Physics of the Earth and Planetary Interiors* 163: 83–105.
Gingold RA and Monaghan JJ (1977) Smoothed particle hydrodynamics: Theory and application to non-spherical stars. *Monthly Notices of the Royal Astronomical Society* 181: 375–389.
Glatzmaier GA (1988) Numerical simulations of mantle convection: Time-dependent, three-dimensional, compressible, spherical shell. *Geophysical and Astrophysical Fluid Dynamics* 43: 223–264.
Glatzmaier GA and Roberts PH (1995) A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* 377: 203–209.
Golub GH and Van Loan CF (1989) *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press, 642 pp.
Gopala VR and van Wachem BGM (2008) Volume of fluid methods for immiscible-fluid and free-surface flows. *Chemical Engineering Journal* 141: 204–221.
Greengard L and Rokhlin V (1987) A fast algorithm for particle simulations. *Journal of Computational Physics* (73).
Griffiths DV and Smith IM (2006) *Numerical Methods for Engineers*, 2nd edn. Chapman and Hall.
Gross L, Bourgouin L, Hale AJ, and Muhlhaus HB (2007) Interface modeling in incompressible media using level sets in Escript. *Physics of the Earth and Planetary Interiors* 163: 23–34.
Gurnis M and Davies GF (1986) Mixing in numerical models of mantle convection incorporating plate kinematics. *Journal of Geophysical Research* 91: 6375–6395.
Gurnis M, Zhong S, and Toth J (2000) On the competing roles of fault reactivation and brittle failure in generating plate tectonics from mantle convection. In: Richards MA, Gordon R, and van der Hilst R (eds.) *The History and Dynamics of Global Plate Motions*, pp. 73–94. Washington, DC: American Geophysical Union.
Hackbush W (1985) *Multigrid Methods and Applications*. Springer.
Hager BH and O'Connell RJ (1981) A simple global model of plate dynamics and mantle convection. *Journal of Geophysical Research* 86: 4843–4878.
Hager BH and Richards MA (1989) Long-wavelength variations in Earth's geoid: Physical models and dynamical implications. *Philosophical Transactions of the Royal Society of London A* 328: 309–327.

Hansen CD and Johnson CR (eds.) (2005) *The Visualization Handbook*. Elsevier, 962 pp.

Hansen U and Yuen DA (1988) Numerical simulations of thermal chemical instabilities and lateral heterogeneities at the core–mantle boundary. *Nature* 334: 237–240.

Harder H (1991) Numerical simulation of thermal convection with Maxwellian viscoelasticity. *Journal of Non-Newtonian Fluid Mechanics* 39: 67–88.

Harder H and Hansen U (2005) A finite-volume solution method for thermal convection and dynamo problems in spherical shells. *Geophysical Journal International* 161: 522–537.

Heroux H, et al. (2013) Trilinos Web page, http://trilinos.sandia.gov.

Hirt CW and Nichols BD (1982) Volume of fluid (VOF) method for the dynamics of free boundaries. *Journal of Computational Physics* 39: 201–225.

Hoover WG and Hoover CG (2001) Smoothed particle hydrodynamics. *Computing in Science and Engineering* 3: 78–85.

Hu J, et al. (2013) ML Web page, http://trilinos.sandia.gov/packages/ml.

Hughes TJR (2000) *The Finite Element Method*. New York: Dover Publications, 682 pp.

Imperial College London (2013) Fluidity Web page, http://www3.imperial.ac.uk/earthscienceandengineering/research/amcg/fluidity.

Ismail-Zadeh A and Tackley P (2010) *Computational Methods for Geodynamics*. Cambridge: Cambridge University Press.

Ita JJ and King SD (1998) The influence of thermodynamic formulation on simulations of subduction zone geometry and history. *Geophysical Research Letters* 25: 1463–1466.

Jadamec MA and Billen MI (2010) Reconciling surface plate motions with rapid three-dimensional mantle flow around a slab edge. *Nature* 465: 338–342. http://dx.doi.org/10.1038/nature09053.

Jadamec MA, Billen MI, and Kreylos O (2012) Three-dimensional simulations of geometrically complex subduction with large viscosity variations. In: *ACM Proceedings of the 2012 Conference on eXtreme Science and Engineering Discovery Environment* http://dx.doi.org/10.1145/2335755.2335827.

Jarvis GT and McKenzie DP (1980) Convection in a compressible fluid with infinite Prandtl number. *Journal of Fluid Mechanics* 96: 515–583.

Jeffers J and Reinders J (2013) *Intel Xeon Phi Coprocessor High Performance Programming*. Morgan Kaufmann.

Kageyama A and Sato T (2004) The "Yin and Yang Grid": An overset grid in spherical geometry. *Geochemistry, Geophysics, Geosystems* 5(9).

Kameyama M, Kageyama A, and Sato T (2008) Multigrid-based simulation code for mantle convection in spherical shell using Yin-Yang grid. *Physics of the Earth and Planetary Interiors* 171: 19–32. http://dx.doi.org/10.1016/j.pepi.2008.06.025.

Kameyama MC and Yuen DA (2006) 3-D convection studies on the thermal state of the lower mantle with post perovskite transition. *Geophysical Research Letters* (33). http://dx.doi.org/10.1029/2006GL025744.

Karato S and Wu P (1993) Rheology of the upper mantle: A synthesis. *Science* 260: 771–778.

Katz R, Knepley MG, Smith B, Spiegelman M, and Coon E (2007) Numerical simulation of geodynamic processes with the Portable Extensible Toolkit for Scientific Computation. *Physics of the Earth and Planetary Interiors* 163: 52–68.

Keller T, May DA, and Kaus BJP (2013) Numerical modelling of magma dynamics coupled to tectonic deformation of lithosphere and crust. *Geophysical Journal International* 195, 1406–1442.

Kellogg LH (1992) Mixing in the mantle. *Annual Review of Earth and Planetary Sciences* 20: 365–388.

Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283: 1881–1884.

King SD and Hager BH (1990) The relationship between plate velocity and trench viscosity in Newtonian and power-law subduction calculations. *Geophysical Research Letters* 17: 2409–2412.

King SD, Lee CY, van Keken PE, et al. (2010) A community benchmark for 2D cartesian compressible convection in the Earth's mantle. *Geophysical Journal International* 180: 73–87. http://dx.doi.org/10.1111/j.1365-246X.2009.04413.x.

King SD, Lowman JP, and Gable CW (2002) Episodic tectonic plate reorganizations driven by mantle convection. *Earth and Planetary Science Letters* 203: 83–91.

King SD, Raefsky A, and Hager BH (1990) ConMan: Vectorizing a finite element code for incompressible two-dimensional convection in the Earth's mantle. *Physics of the Earth and Planetary Interiors* 59: 195–207.

Knepley MG, Brown J, Rupp K, and Smith BF (2013) *Achieving High Performance with Unified Residual Evaluation*, ArXiv e-prints.

Knepley MG and Terrel AR (2013) Finite element integration on GPUs. *ACM Transactions on Mathematical Software* 39(2).

Komatitsch D and Tromp J (1999) Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International* 139: 806–822.

Kowalik Z and Murty TS (1993) *Numerical Modeling of Ocean Dynamics*. Singapore: World Scientific Publishing, 481 pp.

Kreylos O, Bawden GW, Bernardin T, et al. (2006) In: *Enabling Scientific Workflows in Virtual Reality*, *Proceedings of ACM SIGGRAPH*, pp. 155–162.

Kronbichler M, Heister T, and Bangerth W (2012) High accuracy mantle convection simulation through modern numerical methods. *Geophysical Journal International* 191: 12–29.

Kuang W and Bloxham J (1999) Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: Weak and strong field dynamo action. *Journal of Computational Physics* 153: 51–81.

Larsen TB, Yuen DA, Moser J, and Fornberg B (1997) A higher-order finite-difference method applied to large Rayleigh number mantle convection. *Geophysical and Astrophysical Fluid Dynamics* 84: 53–83.

Lenardic A and Kaula WM (1993) A numerical treatment of geodynamic viscous flow problems involving the advection of material interfaces. *Journal of Geophysical Research* 98: 8243–8269.

Leng W and Zhong SJ (2008a) Controls on plume heat flux and plume excess temperature. *Journal of Geophysical Research* 113: B04408. http://dx.doi.org/10.1029/2007JB005155.

Leng W and Zhong SJ (2008b) Viscous heating, adiabatic heating and energetic consistency in compressible mantle convection. *Geophysical Journal International* 173: 693–702.

Leng W and Zhong SJ (2011) Implementation and application of adaptive mesh refinement for thermochemical mantle convection studies. *Geochemistry, Geophysics, Geosystems* 12: Q04006. http://dx.doi.org/10.1029/2010GC003425.

Liao X, Xiao L, Yang C, and Lu Y (2014) Milky Way-2 supercomputer: system and applications, Frontier Compt. *Sci*. 8: 345–356.

Liu X and Zhong SJ (2013) Analysis of marginal stability, heat transfer and boundary layer properties for thermal convection in a compressible fluid. *Geophysical Journal International* 194: 125–144.

Logg A, Olgaard K, Rognes ME, et al. (2011) The FEniCS project, a continually updated technical report at http://fenicsproject.org.

Lynch DR (2005) *Numerical Partial Differential Equations for Environmental Scientists and Engineers: A First Practical Course*. Springer, 388 pp.

Machetel Ph, Rabinowicz M, and Bernadet P (1986) Three-dimensional convection in spherical shells. *Geophysical and Astrophysical Fluid Dynamics* 37: 57–84.

Maday Y and Patera A (1989) Spectral-element methods for the incompressible Navier–Stokes equations. In: Noor A and Oden J (eds.) *State of the Art Survey in Computational Mechanics*, pp. 71–143. New York: ASME.

Malevsky AV (1996) Spline-characteristic method for simulation of convective turbulence. *Journal of Computational Physics* 123: 466–475.

Malevsky AV and Yuen DA (1991) Characteristics-based methods applied to infinite Prandtl number thermal convection in the hard turbulent regime. *Physics of Fluids* A3(no. 9): 2105–2115.

Malevsky AV and Yuen DA (1993) Plume structures in the hard-turbulent regime of three-dimensional infinite Prandtl number convection. *Geophysical Research Letters* 20: 383–386.

May DA and Moresi L (2008) Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics. *Physics of the Earth and Planetary Interiors* http://dx.doi.org/10.1016/j.pepi.2008.07.036.

May DA, Schellart WP, and Moresi LN (2013) Overview of adaptive finite element analysis in computational geodynamics. *Journal of Geodynamics* 70: 1–20.

McKenzie DP, Roberts JM, and Weiss NO (1974) Convection in the Earth's mantle: Towards a numerical solution. *Journal of Fluid Mechanics* 62: 465–538.

McNamara AK and Zhong SJ (2004) Thermochemical structures within a spherical mantle: Superplumes or piles? *Journal of Geophysical Research* 109: B07402. http://dx.doi.org/10.1029/2003JB002847.

Meuer H, Strohmaier E, Dongarra J, and Simon HD (2001) Top500 supercomputer sites. In: *Proceedings of SC*, pp. 10–16.

Monnereau M and Yuen DA (2002) How flat is the lower-mantle temperature gradient? *Earth and Planetary Science Letters* 202: 171–183.

Moresi LN, Dufour F, and Muhlhaus H-B (2003) A Lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials. *Journal of Computational Physics* 184: 476–497.

Moresi L and Gurnis M (1996) Constraints on the lateral strength of slabs from three-dimensional dynamic flow models. *Earth and Planetary Science Letters* 138: 15–28.

Moresi LN, Mühlhaus H-B, Lemiale V, and May D (2007a) Incompressible viscous formulations for deformation and yielding of the lithosphere. In: Karner GD, Manatschal G, and Pinheiro L (eds.). *Imaging, Mapping, and Modeling Extensional*

Processes and Systems, pp. 457–472. The Geological Society. http://dx.doi.org/10.1144/SP282.19, (Special Publication 282).
Moresi LN, Quenette S, Lemiale V, Mériaux C, Appelbe W, and Mühlhaus HB (2007b) Computational approaches to studying non-linear dynamics of the crust and mantle. *Physics of the Earth and Planetary Interiors* 163: 69–82.
Moresi LN and Solomatov VS (1995) Numerical investigation of 2D convection with extremely large viscosity variation. *Physics of Fluids* 9: 2154–2164.
Moresi LN and Solomatov VS (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic styles of the Earth and Venus. *Geophysical Journal International* 133: 669–682.
Moresi LN, Zhong SJ, and Gurnis M (1996) The accuracy of finite element solutions of Stokes' flow with strongly varying viscosity. *Physics of the Earth and Planetary Interiors* 97: 83–94.
Morra G, Regenauer-Lieb K, and Giardini D (2006) Curvature of oceanic arcs. *Geology* 34(10).
Morra G, Yuen DA, Boschi L, Chatelain P, Koumoutsakos P, and Tackley PJ (2010) The fate of the slabs interacting with a density/viscosity hill in the mid-mantle. *Physics of the Earth and Planetary Interiors* 180(3): 271–282.
Muhlhaus H-B and Regenauer-Lieb K (2005) Towards a self-consistent plate mantle model that includes elasticity: Simple benchmarks and application to basic modes of convection. *Geophysical Journal International* 163: 788–800.
Murakami M, Hirose K, Kawamura K, Sata N, and Ohishi Y (2004) Post-perovskite phase transition in $MgSiO_3$. *Science* 304: 855–858.
Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2012) Radial 1-D seismic structures in the deep mantle in mantle convection simulations with self-consistently calculated mineralogy. *Geochemistry, Geophysics, Geosystems* 13: Q11002.
Ogawa M, Schubert G, and Zebib A (1991) Numerical simulations of 3-dimensional thermal convection in a fluid with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 233: 299–328.
Ohno N, Kageyama A, and Kusano K (2006) Virtual reality visualization by CAVE with VFIVE and VTK. *Journal of Plasma Physics* 72: 1069–1072.
Oosterlee CW and Gaspar-Lorenz FJ (2006) Multigrid methods for the Stokes system. *Computing in Science and Engineering* 8(6): 34–43.
OpenFOAM (2013) http://www.openfoam.
Osher S and Fedkiw RP (2003) *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 275 pp.
Pantakar SV (1980) *Numerical Heat Transfer and Fluid Flow*. New York: Hemisphere Publishing.
Parmentier EM and Sotin C (2000) Three-dimensional numerical experiments on thermal convection in a very viscous fluid: Implications for the dynamics of a thermal boundary layer at high Rayleigh number. *Physics of Fluids* 12: 609–617.
Parmentier EM, Turcotte DL, and Torrance KE (1975) Numerical experiments on the structure of mantle plumes. *Journal of Geophysical Research* 80: 4417–4425.
Parmentier EM, Turcotte DL, and Torrance KE (1976) Studies of finite amplitude non-Newtonian thermal convection with application to convection in the Earth's mantle. *Journal of Geophysical Research* 81: 1839–1846.
Patera AT (1984) A spherical element method for fluid dynamics - Laminar flow in a channel expansion. *Journal of Computational Physics* 54: 468–488.
Paulson A, Zhong SJ, and Wahr J (2005) Modelling post-glacial rebound with lateral viscosity variations. *Geophysical Journal International* 163(1): 357–371.
Peltier WR (1976) Glacial isostatic adjustment II. The inverse problem. *Geophysical Journal of the Royal Astronomical Society* 46: 669–705.
Peyret R and Taylor TD (1983) *Computational Methods for Fluid Flow*. Springer, 350 pp.
Phillips JR and White JK (1997) A precorrected-FFT method for electrostatic analysis of complicated 3-D structures. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 16: 1059–1072.
Poulson J, Marker B, Hammond JR, Romero NA, and van de Geijn R (2013) Elemental: A new framework for distributed memory dense matrix computations. *ACM Transactions on Mathematical Software* 39(2).
Press WH, Flannery BP, Teukolsky SA, and Vetterling WA (1992) *Numerical Recipes in Fortran*.
Ramage A and Wathen AJ (1994) Iterative solution techniques for the Stokes and Navier–Stokes equations. *International Journal for Numerical Methods in Fluids* 19: 67–83.
Ratcliff JT, Schubert G, and Zebib A (1996) Steady tetrahedral and cubic patterns of spherical-shell convection with temperature-dependent viscosity. *Journal of Geophysical Research* 101(25): 25473–25484.
Richter FM (1973) Finite amplitude convection through a phase boundary. *Geophysical Journal of the Royal Astronomical Society* 35: 265–276.
Richtmeyer RD and Morton KW (1967) *Difference Methods for Initial Value Problems*. Interscience Publishers, 405pp.
Rupp K, Rudolf F, and Weinbub J (2010) ViennaCL—A high level linear algebra library for GPUs and multi-core CPUs. *International Workshop on GPUs and Scientific Applications* (51–56).
Saad Y and Schultz MH (1986) GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific Computing* 7: 856–869.
Schmalholz SM, Podladchikov YY, and Schmid DW (2001) A spectral/finite-difference for simulating large deformations of heterogeneous, viscoelastic materials. *Geophysical Journal International* 145: 199–219.
Schott B and Schmeling H (1998) Delamination and detachment of a lithospheric root. *Tectonophysics* 296: 225–247.
Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge University Press.
Schubert G, Yuen DA, and Turcotte DL (1975) Role of phase transitions in a dynamic mantle. *Geophysical Journal of the Royal Astronomical Society* 42: 705–735.
Smolarkiewicz PK (1983) A simple positive definite advection transport algorithm with small implicit diffusion. *Monthly Weather Review* 111: 479–489.
Smolarkiewicz PK (1984) A fully multidimensional positive definite advection scheme with small implicit diffusion. *Journal of Computational Physics* 54: 325–339.
Spalding DB (1972) A novel finite difference formulation for differential expressions involving both first and second derivatives. *International Journal for Numerical Methods in Engineering* 4: 551–559.
Stadler G, Gurnis M, Burstedde C, Wilcox LC, Alisic L, and Ghattas O (2010) The dynamics of plate tectonics and mantle flow: From local to global scales. *Science* 329: 1033–1038.
Stemmer K, Harder H, and Hansen U (2006) A new method to simulate convection with strongly temperature-dependent and pressure-dependent viscosity in a spherical shell: Applications to the Earth's mantle. *Physics of the Earth and Planetary Interiors* 157: 223–249.
Sulsky D, Chen Z, and Schreyer HL (1994) A particle method for history-dependent materials. *Computer Methods in Applied Mechanics and Engineering* 118: 179–196.
Tackley PJ (1993) Effects of strongly temperature-dependent viscosity on time-dependent, 3-dimensional models of mantle convection. *Geophysical Research Letters* 20: 2187–2190.
Tackley PJ (1994) *Three-Dimensional Models of Mantle Convection: Influence of Phase Transitions and Temperature-Dependent Viscosity*. PhD Thesis, Pasadena, CA: California Institute of Technology.
Tackley PJ (1998a) Three-dimensional simulations of mantle convection with a thermo-chemical basal boundary layer: D″. In: Gurnis M (ed.) *The Core–Mantle Region*, pp. 231–253. Washington, DC: AGU.
Tackley PJ (1998b) Self-consistent generation of tectonic plates in three-dimensional mantle convection. *Earth and Planetary Science Letters* 157: 9–22.
Tackley PJ (2000) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations. 1. Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* (1).
Tackley PJ (2008) Modeling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid. *Physics of the Earth and Planetary Interiors* 171: 7–18. http://dx.doi.org/10.1016/j.pepi.2008.08.005.
Tackley PJ and King SD (2003) Testing the tracer ratio method for modeling active compositional fields in mantle convection simulations. *Geochemistry, Geophysics, Geosystems* (4). http://dx.doi.org/10.1029/2001GC000214.
Tackley PJ, Stevenson DJ, Glatzmaier GA, and Schubert G (1993) Effects of an endothermic phase transition at 670 km depth on spherical mantle convection. *Nature* 361: 137–160.
Tan E and Gurnis M (2005) Metastable superplumes and mantle compressibility. *Geophysical Research Letters* 32: L20307. http://dx.doi.org/10.1029/2005GL024190.
Tan E and Gurnis M (2007) Compressible thermochemical convection and application to lower mantle structures. *Journal of Geophysical Research* 112: B06304. http://dx.doi.org/10.1029/2006JB004505.
Tan E, Leng W, Zhong SJ, and Gurnis M (2011) On the location of plumes and lateral movement of thermo-chemical structures with high bulk modulus in the 3-D compressible mantle. *Geochemistry, Geophysics, Geosystems* 12: Q07005. http://dx.doi.org/10.1029/2011GC003665.
Torrance KE and Turcotte DL (1971) Thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 47: 113–125.
Travis BJ, Olson P, and Schubert G (1990) The transition from two-dimensional to three-dimensional planforms in infinite-Prandtl-number thermal convection. *Journal of Fluid Mechanics* 216: 71–92.
Trompert RA and Hansen U (1996) The application of a finite-volume multigrid method to 3-dimensional flow problems in a highly viscous fluid with a variable viscosity. *Geophysical and Astrophysical Fluid Dynamics* 83: 261–291.

Trompert RA and Hansen U (1998) Mantle convection simulations with rheologies that generate plate-like behavior. *Nature* 395: 686–688.

Trottenberg U, Oosterlee C, and Schueller A (2001) *Multigrid*. Academic Press, 631 pp.

Turcotte DL, Torrance KE, and Hsui AT (1973) Convection in the Earth's mantle. *Methods*. *Computational Physics*, vol. 13, pp. 431–454. Academic Press.

Valencia D, Sasselov DD, and O'Connell RJ (2007) Radius and structure models of the first super-Earth planet. *Astrophysical Journal* 656: 545–551.

van den Berg AP, van Keken PE, and Yuen DA (1993) The effects of a composite non-Newtonian and Newtonian rheology on mantle convection. *Geophysical Journal International* 115: 62–78.

van Heck H and Tackley PJ (2008) Planforms of self-consistently generated plate tectonics in 3-D spherical geometry. *Geophysical Research Letters* 35: L19312. http://dx.doi.org/10.1029/2008GL035190.

van Keken PE, King SD, Schmeling H, Christensen UR, Neumeister D, and Doin M-P (1997) A comparison of methods for the modeling of thermochemical convection. *Journal of Geophysical Research* 102: 22477–22496.

van Keken PE and Zhong S (1999) Mixing in a 3D spherical model of present-day mantle convection. *Earth and Planetary Science Letters* 171: 533–547.

Washio T and Oosterlee CW (2000) Krylov subspace acceleration for nonlinear multigrid schemes with application to recirculating flow. *SIAM Journal on Scientific Computing* 21: 1670–1690.

Webster DR and Longmire EK (2001) Jet pinch off and drop formation in immiscible liquid–liquid system. *Experiments in Fluids* 30: 47–56.

Weinberg RF and Schmeling H (1992) Polydiapirs: Multi-wavelength gravity structures. *Journal of Structural Geology* 14: 425–436.

Wells G, Logg A, and Mardal K-A (eds.) (2012) *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book*. In: *Lecture Notes in Computational Science and Engineering*, vol. 84. Springer.

Whitehead JA and Luther DS (1975) Dynamics of laboratory diapir and plume models. *Journal of Geophysical Research* 80(5): 705–717.

Wright GB, Flyer N, and Yuen D (2010) A hybrid radial basis function—Pseudospectral method for thermal convection in a 3{D} spherical shell. *Geochemistry, Geophysics, Geosystems,* Q07003.

Yavneh I (2006) Why multigrid methods are so efficient? *Computing in Science and Engineering* 8: 12–23.

Ying L, Biros G, Zorin D, and Langston H (2003) A new parallel kernel-independent fast multipole method. In: *16th ACM/IEEE Conference on Supercomputing*.

Yoshida M and Kageyama A (2004) Application of the Yin-Yang grid to a thermal convection of a Boussinesq fluid with infinite Prandtl number in a three-dimensional spherical shell. *Geophysical Research Letters* 31. http://dx.doi.org/10.1029/2004GL019970.

Yuen DA, Balachandar S, and Hansen U (2000) Modelling mantle convection: A significant challenge in geophysical fluid dynamics. In: Fox PA and Kerr RM (eds.) *Geophysical and Astrophysical Convection*, pp. 257–294.

Yuen DA, Wang L, Chi X, Ge W, Johnsson L, and Shi Y (2013) *GPU Solutions to Multiscale Problems in Science and Engineering*. Springer.

Zhang S and Christensen U (1993) Some effects of lateral viscosity variations on geoid and surface velocities induced by density anomalies in the mantle. *Geophysical Journal International* 114: 531–547.

Zhang S and Yuen DA (1995) The influences of lower mantle viscosity stratification on 3-D spherical-shell mantle convection. *Earth and Planetary Science Letters* 132: 157–166.

Zhang S and Yuen DA (1996a) Various influences on plumes and dynamics in time-dependent, compressible, mantle convection in 3-D spherical shell. *Physics of the Earth and Planetary Interiors* 94: 241–267.

Zhang S and Yuen DA (1996b) Intense local toroidal motion generated by variable viscosity compressible convection in 3-D spherical shell. *Geophysical Research Letters* 23: 3135–3138.

Zhong SJ (2006) Constraints on thermochemical convection of the mantle from plume heat flux, plume excess temperature and upper mantle temperature. *Journal of Geophysical Research* 111: B04409. http://dx.doi.org/10.1029/2005JB003972.

Zhong SJ and Gurnis M (1994) The role of plates and temperature-dependent viscosity in phase change dynamics. *Journal of Geophysical Research* 99: 15903–15917.

Zhong SJ, Gurnis M, and Moresi LN (1998) The role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models. *Journal of Geophysical Research* 103: 15255–15268.

Zhong SJ and Hager BH (2003) Entrainment of a dense layer by thermal plumes. *Geophysical Journal International* 154: 666–676.

Zhong SJ, McNamara A, Tan E, Moresi L, and Gurnis M (2008) A benchmark study on mantle convection in a 3-D spherical shell using CitcomS. *Geochemistry, Geophysics, Geosystems* (9). http://dx.doi.org/10.1029/2008GC002048.

Zhong SJ, Paulson A, and Wahr J (2003) Three-dimensional finite-element modelling of Earth's viscoelastic deformation: Effects of lateral variations in lithospheric thickness. *Geophysical Journal International* 155(2): 679–695.

Zhong SJ, Zuber MT, Moresi LN, and Gurnis M (2000) Role of temperature dependent viscosity and surface plates in spherical shell models of mantle convection. *Journal of Geophysical Research* 105: 11063–11082.

Zienkiewicz OC, Taylor RL, and Zhu JZ (2005) *The Finite Element Method: Its Basis and Fundamentals*, 6th edn. Butterworth and Heinemann.

7.06 Temperatures, Heat, and Energy in the Mantle of the Earth

C Jaupart, Institut de Physique du Globe de Paris, Paris, France
S Labrosse, Ecole Normale Supérieure de Lyon, Lyon, France
F Lucazeau, Institut de Physique du Globe de Paris, Paris, France
J-C Mareschal, Université du Québec à Montréal, Montréal, QC, Canada

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by C Jaupart, S Labrosse and J.-C Mareschal, Volume 7, pp. 253–303, © 2007, Elsevier B.V.

7.06.1 Introduction 224
7.06.2 Basic Thermodynamics 225
7.06.2.1 Breakdown of the Energy Budget 225
7.06.2.2 Changes in Gravitational Energy: Contraction due to Secular Cooling 226
7.06.2.3 Secular Cooling Equation 227
7.06.2.4 Summary 228
7.06.3 Heat Loss Through the Seafloor 228
7.06.3.1 Oceanic Heat Flow Measurements 228
7.06.3.2 Cooling of the Oceanic Lithosphere 229
7.06.3.2.1 The shallow oceanic environment 230
7.06.3.2.2 Initial condition: Temperature distribution at the ridge axis 230
7.06.3.2.3 Bottom boundary condition 230
7.06.3.3 Validating the Cooling Model for Oceanic Lithosphere 231
7.06.3.3.1 High-resolution surveys near mid-ocean ridges 231
7.06.3.3.2 Global analysis 232
7.06.3.3.3 Depth of the seafloor 233
7.06.3.4 Heat Loss Through the Ocean Floor 234
7.06.3.4.1 Global oceanic heat loss calculation 234
7.06.3.4.2 Heat loss through young ocean floor 235
7.06.3.4.3 Hot spots 236
7.06.3.5 Summary 236
7.06.4 Heat Loss Through Continents 236
7.06.4.1 Average Continental Heat flux and Heat Loss Through Continental Areas 236
7.06.4.2 Various Contributions to the Surface Heat Flux in Continental Areas 237
7.06.4.3 Estimating Moho Heat Flux 237
7.06.4.3.1 Relationship between heat flow and heat production 237
7.06.4.3.2 Crustal heat production and Moho heat flux 238
7.06.4.4 Recently Active Regions and Continental Margins 239
7.06.4.4.1 Compressional orogens 239
7.06.4.4.2 Rifts and continental zones of extension 240
7.06.4.4.3 Continental margins: Passive 240
7.06.4.4.4 Continental margins: Active 242
7.06.4.5 Mantle Heat Loss Through Continental Areas 242
7.06.4.6 Summary 243
7.06.5 Heat Sources 243
7.06.5.1 Radiogenic Sources in the Bulk Silicate Earth 243
7.06.5.2 Geoneutrino Measurements 245
7.06.5.3 Heat Flux from the Core 246
7.06.5.4 Other Sources: Tidal Heating and Crust–Mantle Differentiation 248
7.06.5.5 Summary 248
7.06.6 Secular Cooling: Constraints on Mantle Temperatures 248
7.06.6.1 The Present-Day Mantle Geotherm 249
7.06.6.2 Temperature Versus Time 251
7.06.6.3 Early Earth 251
7.06.6.4 Magma Ocean Evolution 252
7.06.6.5 Average Secular Cooling Rate of the Mantle 253
7.06.6.6 Summary 254
7.06.7 Thermal Evolution Models 254
7.06.7.1 The Urey Ratio 254
7.06.7.2 'Parameterized' Cooling Models 254
7.06.7.3 The Peculiarities of Mantle Convection: Observations 255

7.06.7.4 Convection with Oceanic Plates 256
7.06.7.5 Vagaries of Seafloor Spreading and Heat Loss 256
7.06.7.6 Heat Flow from the Core 257
7.06.7.7 Summary 258
7.06.8 Conclusions 258
Appendix A Contraction of the Earth due to Secular Cooling 258
Appendix B Gravitational Energy Changes 259
Appendix C Viscous Dissipation 259
Appendix D Half-Space Cooling Model with Temperature-Dependent Properties 261
Appendix E Plate Models for the Oceanic Lithosphere 261
Appendix F Differences Between Estimates of the Energy Budget 261
Appendix G Average Thermal Structure and Temperature Changes in Upwellings and Downwellings 262
Appendix H Seafloor Age Distribution as Seen from Models of Mantle Convection 263
Acknowledgments 264
References 264

7.06.1 Introduction

Even a cursory historical look at the development of geology reveals that the cooling of the Earth has always been a central issue (Birch, 1965; Fourier, 1824; Holmes, 1915; Joly, 1909; McDonald, 1959; Strutt, 1906; Thomson, 1864; Urey, 1964). From Hutton's view that there is no beginning or end to geologic time to today's debate on how far back in time plate tectonics can be traced, arguments eventually boil down to statements on our planet's thermal evolution. At the time of Fourier and Kelvin, the Earth's temperature gradient was estimated to be in the range 20–30 K km^{-1}, close to what we know today after more than six decades of heat flow measurements on land and at sea. Kelvin thought that this piece of data would constrain the age of the Earth and concluded from his famous calculation that the Earth was probably not much older than 100 My. The failure of Kelvin's model has usually been attributed to two assumptions, namely, that the Earth is cooling by conduction and that there are no sources of heat inside the Earth. In retrospect, one could argue that an equally damaging flaw was the inadequate data set that was available to him. Had Kelvin known that the heat flux varies by large amounts at the Earth's surface, he might not have advocated a simple uniform cooling model for the whole planet. The large lateral variations of heat flux that occur on Earth provide information on cooling mechanisms and heat sources that are as important as the global average.

Today, we know that the Earth is not cooling by conduction only and that the convective motions that drive the oceanic plates are responsible for a large fraction of its heat loss. We also know that heat generated by the radioactive decay of U, Th, and K in silicate rocks accounts for a large fraction of the surface heat loss. It is fair to state, however, that the Earth's heat engine works in ways that still elude us. In this context, it is worth remembering that we pursue two different goals in studies of mantle convection. One is to account for its present-day dynamics and for its role in shaping the Earth's surface features. The other goal is to go back in time in order to evaluate how the rates of geologic processes have changed and to decipher processes that are no longer active today. Both goals require a thorough understanding of the Earth's heat budget, but each relies on a different set of constraints. The former can be attained using present-day observations, such as the distributions of seismic velocities and density at depth. The latter goal requires knowledge of how the convective regime evolved as the Earth cooled down and its energy sources ran down. From another standpoint, the present-day energy budget of the planet and the distribution of heat flux at the surface are constraints that mantle convection models must satisfy.

The present-day energy budget reflects how the Earth's convective engine has evolved through geologic time. The power of this constraint has motivated a large number of studies. With hindsight, one may link the emergence of convection models to the failure of conductive (and radiative) thermal history models to account for the mantle temperature regime and the Earth energy budget (Jacobs, 1961; Jacobs and Allan, 1956; MacDonald, 1959). Convection in the Earth's mantle had become inescapable and was required to explain why the mantle is not fully molten today. The difficulty in running fully consistent dynamic calculations of convection over the whole history of our planet led to the so-called parameterized models such that the heat flux out of the Earth is written directly as a function of dimensionless numbers describing the bulk convective system, such as the Rayleigh number (Schubert et al., 2001). For a given set of initial conditions, the model results were required to match the present-day energy budget or more precisely the ratio of heat production to heat loss (the 'Urey' ratio). The earliest study of this kind was probably that of McKenzie and Weiss (1975) and was followed by countless others. This approach was used to argue against whole-layer mantle convection (McKenzie and Richter, 1981), to date the emergence of plate tectonics (Peltier and Jarvis, 1982), to derive constraints on the distribution of radiogenic heat sources in the mantle (Schubert et al., 1980), and even to determine the amount of radioactive sources in the Earth's core (Breuer and Spohn, 1993; Davies, 1980a,b). The difficulty in accounting for the wealth of processes that characterize the Earth, such as continental growth as well as degassing and the implied changes of rheological properties, however, has led to disenchantment. Yet, it is clear that the present-day thermal and tectonic regime of the Earth results from several billion years of convective processes and is best understood within a time-dependent framework.

Determination of the Earth's rate of heat loss requires a very large number of heat flux measurements in a variety of geologic settings. Local surveys as well as global analyses of large data sets have shown that heat flux varies on a wide range of spatial

scales and, in the continents, is not a function of a single variable such as geologic age. Heat flux data exhibit large scatter, which has had unfortunate consequences. One has been that few scientists have invested time and energy to sort out the large number of physical processes that come into play. The 1980s saw a rapid decrease in the number of research teams active in that field as well as in the number of measurements carried out at sea and on land. Another consequence has been that, with few notable exceptions (e.g., Coltice et al., 2012; Pari and Peltier, 2000), the distribution of heat flux is rarely used to test the validity of mantle convection calculations. Many evolutionary models for the Earth's mantle have abandoned the energy budget as a viable constraint and rely on geochemical data. Here, we will assess the reliability of heat flux measurements and shall demonstrate that the spatial distribution of heat flux provides a key constraint for understanding convection in the Earth.

The last three decades have seen notable advances in the interpretation of heat flux. In the oceans, these include a thorough understanding of hydrothermal circulation through oceanic crust and sediments, as well as detailed and precise heat flux measurements through both very young and very old seafloor. In the continents, sampling of old cratons is now adequate in several areas, heat production of lower crustal assemblages is better understood, and systematic studies of heat flux and heat production allow strong constraints on the crustal contribution to the surface heat flux. Today, we have a better understanding of the energy sources in the Earth than we did 30 years ago and know how large some of the uncertainties are.

In this chapter, we shall focus on two different, but closely related, problems. One is to evaluate the present-day energy budget of the mantle with emphasis on the associated uncertainties. The other is to evaluate how thermal evolution models must be developed in order to account for this budget. We shall establish the gross thermodynamics of the Earth and shall explain how estimates of heat loss and heat production have been obtained, drawing from recent advances. We shall emphasize the peculiarities of heat loss mechanisms of our planet and in particular the spatial distribution of heat flux. We shall then rely on this budget to infer the present-day secular cooling rate of our planet. We shall also evaluate independent constraints on temperature in the Earth's mantle and present a reference temperature profile through the convective mantle. Finally, we shall discuss the thermal evolution of our planet, from the standpoint of both observations and theoretical models. In order to facilitate the reader's task, we give a short summary of major points at the end of each section.

7.06.2 Basic Thermodynamics

7.06.2.1 Breakdown of the Energy Budget

The integral form of the energy balance for the whole planet takes the form

$$\frac{\mathrm{d}\left(U+E_\mathrm{c}+E_\mathrm{g}\right)}{\mathrm{d}t}=-\int_\mathrm{A}\boldsymbol{q}\cdot\boldsymbol{n}\,\mathrm{d}A+\int_\mathrm{V}H\mathrm{d}V+\int_\mathrm{V}\psi\mathrm{d}V-p_\mathrm{a}\frac{\mathrm{d}V}{\mathrm{d}t} \quad [1]$$

where U is internal energy, E_c is kinetic energy, and E_g is gravitational potential energy. $\boldsymbol{q}$ is the surface heat flux, n is the unit normal vector, A is the Earth's outer surface, H is internal heat production per unit volume, p_a is atmospheric pressure, V is the Earth's total volume, and ψ is energy transfers to or from external systems, such as tidal dissipation. Table 1 provides a list of the main symbols used in this chapter. Equation [1] states that the Earth's total energy changes due to heat loss, internal heat generation, energy transfer between our planet's interior and its surroundings (atmosphere as well as other celestial bodies), and finally the work of atmospheric pressure as the planet contracts. We have assumed that the Earth's surface is shear stress-free. Dissipation induced by internal convective motions and chemical differentiation is not written in because it is involved in internal transfers of energy and does not act to change the total energy of the system. The state variables include pressure, temperature, and chemical composition, which may all vary as a function of time and

Table 1 Symbols used

Symbol	*Definition*	*Units (commonly used units or value)*
C_p	Heat capacity	J kg^{-1} K^{-1}
C_Q	Heat flux/age$^{1/2}$	470 – 510 mW m^{-2}/(My)$^{1/2}$
C_A	Seafloor accretion rate	3.34 km^2 My^{-1}
D	Thickness of convecting layer	m
E_c	Kinetic energy	J
E_g	Gravitational potential energy	J
E_rot	Rotational energy	J
E_s	Strain energy	J
F_b	Buoyancy flux	W
g	Acceleration of gravity	m s^{-1}
G	Gravitational constant	6.67×10^{-11}kg m^3 s^{-2}
H	(Volumetric) heat production	W m^{-3} (μW m^{-3})
I	Moment of inertia	kg m^2
K	Bulk modulus	Pa
L	Length of oceanic plate (length scale)	m
M	Mass of Earth	kg
Q, q	Heat flux	W m^{-2} (mW m^{-2})
p	Pressure	Pa (MPa, GPa)
R	Radius of Earth	km (6378)
Ra	Rayleigh number	/
s	Entropy per unit mass	J kg^{-1} K^{-1}
T	Temperature	K (°C)
u	Internal energy per unit mass (also used for horizontal velocity component)	J kg^{-1}
U	Internal energy (also used for horizontal velocity component)	J
Ur	Urey number	
V	Volume	m^3
w	Radial velocity	
α	Volumetric expansion coefficient	K^{-1}
γ	Grüneisen parameter	
κ	Thermal diffusivity	m^2 s^{-1}
λ	Thermal conductivity	W m^{-1} K^{-1}
μ	Viscosity	Pa s
Ω	Angular velocity	rad s^{-1}
Φ	Heat dissipated by friction	W m^{-3}
Ψ	External energy sources	W m^{-3}
ρ	Density	kg m^{-3}
ΔS_cond	Entropy production	J K^{-1}
σ	Deviatoric stress tensor	Pa (MPa)

position within the planet. Chemical differentiation proceeds mostly by the gravity-driven separation of different phases, which entails the transformation of gravitational potential energy into heat by viscous dissipation with no change of the total energy. In this chapter, we shall not dwell at length on such processes. The largest segregation events, such as the formation of the Earth's core and, possibly, the separation of two distinct mantle reservoirs through crystal–melt separation in a magma ocean, occurred very early on. These events set up the initial conditions for secular thermal evolution involving no major chemical separation process energy-wise. Extraction of continental crust out of the mantle has been active more or less continuously since the beginning, but we shall show that the magnitude of the induced energy change is small compared to other items in the energy budget. We shall therefore write thermodynamic principles using only temperature and pressure as variables of state. We shall not deal with the orbital energy of the Sun–Earth system and shall not discuss changes of the kinetic energy of the Earth's revolution around the Sun and associated changes of gravitational energy due to variations of the Earth's orbit.

Our main purpose in this chapter is to evaluate the different terms in the energy balance and to derive an equation for the average temperature of the Earth. The dominant terms on the right of the energy balance equation [1] are the Earth's rate of heat loss and internal heat generation, which are inferred from field measurements and chemical Earth models. The other terms are evaluated theoretically and are shown to be negligible. We explain how internal energy transfers operate and show that changes of gravitational energy are compensated by changes of strain energy E_S, which is the energy required to compress matter to its actual local pressure p.

The gravitational energy of the Earth is defined as the energy required to bring matter from infinity and, assuming spherical symmetry, can be written as

$$E_g = -\int_0^R \rho(r) g(r) r 4\pi r^2 \mathrm{d}r \qquad [2]$$

with ρ and g the spatially varying density and gravity. This energy is negative because the accretion process releases energy. This energy can be computed for the present Earth and an upper limit can be obtained for a sphere with uniform density:

$$E_g = -\frac{3}{5}\frac{GM^2}{R} \qquad [3]$$

where G is the gravitational constant and M is the mass of the Earth.

Kinetic energy may be broken down into several different components:

$$E_c = E_{rot} + E_{contr} + E_{conv} \qquad [4]$$

corresponding to the bulk rotation of our planet, radial contraction induced by secular cooling, and internal convective motions, respectively. One may easily show that the latter two are small compared to the first one.

Table 2 lists estimates for gravitational, kinetic, and internal energy components and makes it clear that kinetic energy is very small compared to the other two. A striking result is that the largest component by far is gravitational energy, which is larger than internal energy by at least one order of magnitude. In a constant mass planet, gravitational energy changes due to thermal contraction, chemical differentiation, and vertical movements of the Earth's surface (tectonic processes and erosion–deposition). These various processes work in different ways and are associated with different energy transport mechanisms and hence must be dealt with separately.

Table 2 Numbers: order of magnitude

	Value	*Units*
Rotational energy	2.1×10^{29}	J
Internal energy (for 2500 K average temperature)	1.7×10^{31}	J
Gravitational energy (uniform sphere)	2.2×10^{32}	J
Rotation angular velocity	7.292×10^{-5}	rad s^{-1}
Polar moment of inertia	8.0363×10^{37}	kg m^2
Total mass	5.9737×10^{24}	kg
Total volume	1.08×10^{21}	m^3
Mass mantle	$\approx 4.0 \times 10^{24}$	kg
Mass crust	$\approx 2.8 \times 10^{22}$	kg

7.06.2.2 Changes in Gravitational Energy: Contraction due to Secular Cooling

Gravitational energy is the largest component of the budget (Table 2), and special care is warranted to evaluate how it gets converted to other forms of energy when the planet contracts. This has been discussed in a series of papers (Flasar and Birch, 1973; Lapwood, 1952). Here, we avoid detailed calculations and throw light on some interesting thermodynamic aspects.

The gravitational energy changes when the planet contracts:

$$\frac{\Delta E_g}{E_g} = -\frac{\Delta R}{R} \qquad [5]$$

Neglecting changes of pressure and gravity and assuming uniform cooling by an amount ΔT,

$$\frac{\Delta R}{R} \approx \frac{\langle\alpha\rangle \Delta T}{3} \qquad [6]$$

where $\langle\alpha\rangle$ is an average value for the coefficient of thermal expansion and ΔT is negative. In Appendix A, we show that the changes of pressure and gravity that are induced by the contraction process enhance the contraction rate. Assuming $\langle\alpha\rangle \approx 2 \times 10^{-5}\,\mathrm{K}^{-1}$ and a secular cooling rate of 100 K Gy^{-1}, the contraction velocity is $\mathrm{d}R/\mathrm{d}t \approx -10^{-13}\,\mathrm{m\,s}^{-1}$, a very small value compared to the typical convective velocity of $\approx 10^{-9}$ m s^{-1}. The induced change of gravitational energy, however, is far from being negligible. For the same choice of parameter values, it is ≈ 4 TW, which, as we shall see, corresponds to 10% of the total energy loss of our planet. We shall demonstrate, however, that such changes of gravitational energy are not converted to heat.

Thermal contraction affects the planet's rotation. The moment of inertia I changes:

$$\Delta I/I = 2\Delta R/R \qquad [7]$$

and hence,

$$\Delta E_{rot}/E_{rot} = \Delta\Omega/\Omega = -2\Delta R/R \qquad [8]$$

where Ω is the Earth rotation velocity. Thus, some of the gravitational potential energy goes into the energy of rotation. Rotational energy is much less (three orders of magnitude)

than gravitational energy, however, and hence may be neglected in the present analysis.

To elucidate energy transfer processes, we now consider thermodynamics at the local scale. We focus on a few specific aspects of interest and refer the reader to the study by Braginsky and Roberts (1995) for a comprehensive analysis. All energies are now written per unit mass with small letters, that is, e_c and u stand for the kinetic energy and internal energy per unit mass, respectively. We begin with the standard form of the first law of thermodynamics (Bird et al., 1960; Schubert et al., 2001; see also **Chapter 7.02** by Ricard in this volume):

$$\rho\frac{\mathrm{D}(u+e_c)}{\mathrm{D}t}=-\nabla\cdot\boldsymbol{q}-\nabla\cdot(p\mathbf{v})-\nabla\cdot[\boldsymbol{\sigma}\cdot\mathbf{v}]+H+\psi+\rho\mathbf{g}\cdot\mathbf{v} \qquad [9]$$

where $\boldsymbol{\sigma}$ is the deviatoric stress tensor, $\mathbf{v}$ is velocity, and ψ collects external source terms such as tidal dissipation. We have adopted the convention that $\boldsymbol{\sigma}$ denotes stresses acting on material located on the positive side of a surface. From the momentum equation, we get

$$\rho\frac{\mathrm{D}e_c}{\mathrm{D}t}=\rho\frac{\mathrm{D}\left(\frac{\mathbf{v}^2}{2}\right)}{\mathrm{D}t}=-\mathbf{v}\cdot\boldsymbol{\nabla}p-\mathbf{v}\cdot\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}+\rho\mathbf{g}\cdot\mathbf{v} \qquad [10]$$

Subtracting this from the total energy balance leads to an equation for the internal energy:

$$\rho\frac{\mathrm{D}u}{\mathrm{D}t}=-\nabla\cdot\mathbf{q}+H+\psi+\phi-p\boldsymbol{\nabla}\cdot\mathbf{v} \qquad [11]$$

where ϕ stands for viscous dissipation:

$$\phi=-\nabla\cdot[\boldsymbol{\sigma}\cdot\mathbf{v}]+\mathbf{v}\cdot\nabla\cdot\boldsymbol{\sigma}=-\boldsymbol{\sigma}:\nabla\mathbf{v} \qquad [12]$$

Equation [11] is thus the usual statement that changes of internal energy u are due to heat gains or losses (which are broken into four contributions) and to the work of pressure (the last term on the right).

All the equations earlier stem from standard thermodynamics theory. We now introduce gravitational energy and strain energy. We decompose variables into the sum of the azimuthal average and a perturbation, such as

$$T=\overline{T}+\theta \qquad [13]$$

for temperature. The velocity field is decomposed into a component due to contraction, $\mathbf{v}_c$, and a convective component, $\mathbf{w}$. We consider separately the effects of contraction, which act on the average density and pressure, and the effects of convection, which involve departures from these averages.

We may assume that contraction proceeds in conditions close to hydrostatic equilibrium, such that it involves no deviatoric stress and no dissipation. In this case, the azimuthal average of the momentum equation reduces to a hydrostatic balance:

$$0=-\nabla\overline{p}+\overline{\rho}\mathbf{g} \qquad [14]$$

For the contraction process with velocity $\mathbf{v}_c$, we may break down internal energy changes as follows:

$$\overline{\rho}\frac{\mathrm{D}u}{\mathrm{D}t}=\overline{\rho}T\frac{\mathrm{D}s}{\mathrm{D}t}-\overline{p}\nabla\cdot\mathbf{v}_c \qquad [15]$$

where s is the entropy per unit mass and where we identify the work done by pressure as a change of strain energy:

$$-\overline{p}\nabla\cdot\mathbf{v}_c=\overline{\rho}\frac{\mathrm{D}e_s}{\mathrm{D}t} \qquad [16]$$

In the total energy balance equation [9], the last term on the right is the work done by the gravity force. By definition, this term can be written as the change of gravitational potential energy when it is carried over to the left-hand side of the balance:

$$\overline{\rho}\frac{\mathrm{D}e_g}{\mathrm{D}t}=-\overline{\rho}\mathbf{g}\cdot\mathbf{v}_c \qquad [17]$$

This relationship is demonstrated in integral form in **Appendix B**. Collecting all terms, the energy balance equation [9] is written as

$$\overline{\rho}T\frac{\mathrm{D}s}{\mathrm{D}t}+\overline{\rho}\frac{\mathrm{D}\left[e_s+e_g+e_c\right]}{\mathrm{D}t}=-\nabla\cdot\mathbf{q}+H+\psi-\nabla\cdot(\overline{p}\mathbf{v}_c)+(\cdots) \qquad [18]$$

In this equation, terms associated with convective motions are not written explicitly and will be dealt with later. Kinetic energy is also negligible and, by inspection, one may deduce from eqn [18] that

$$\overline{\rho}\frac{\mathrm{D}\left[e_s+e_g\right]}{\mathrm{D}t}=-\nabla\cdot(\overline{p}\mathbf{v}_c) \qquad [19]$$

This can be demonstrated by recalling the identity

$$\nabla\cdot(\overline{p}\mathbf{v}_c)=\overline{p}\nabla\cdot\mathbf{v}_c+\mathbf{v}_c\cdot\nabla\overline{p} \qquad [20]$$

Using the hydrostatic balance in eqn [17], the right-hand side of this equation can be recast as

$$\overline{p}\nabla\cdot\mathbf{v}_c+\mathbf{v}_c\cdot\nabla\overline{p}=-\overline{\rho}\frac{\mathrm{D}e_s}{\mathrm{D}t}-\overline{\rho}\frac{\mathrm{D}e_g}{\mathrm{D}t} \qquad [21]$$

which is indeed eqn [19].

Integrating eqn [19] over the whole planet volume, we finally obtain

$$\frac{\mathrm{d}E_g}{\mathrm{d}t}+\frac{\mathrm{d}E_s}{\mathrm{d}t}=-p_a\frac{\mathrm{d}V}{\mathrm{d}t} \qquad [22]$$

where E_s is the total strain energy of the planet. The term on the right-hand side is negligible, and this equation therefore states that the change of gravitational energy is compensated by one of strain energy and does not generate heat. Contraction does lead to a small temperature increase, however, as will be shown in the succeeding text.

7.06.2.3 Secular Cooling Equation

To derive an equation for temperature, we need an expression for dissipation and lengthy manipulations that are given in **Appendix C**. Dissipation is balanced by the work of buoyancy forces driving convection (called the 'buoyancy flux') and disappears from the final budget, as expected. This leads to the following temperature equation:

$$\int_V\overline{\rho}C_v\frac{\partial\overline{T}}{\partial t}\mathrm{d}V=\int_V\frac{\alpha TK_T}{\rho}\frac{\partial\rho}{\partial t}\mathrm{d}V-\int_A\mathbf{q}\cdot\mathrm{d}A+\int_V H\mathrm{d}V \qquad [23]$$

where C_v is thermal capacity at constant volume. The first term on the right is the only remaining contribution of contraction and has been called 'adiabatic' heating. It can be estimated as follows:

$$E_a = \int_V \frac{\alpha T K_T}{\rho}\frac{\partial \rho}{\partial t}\mathrm{d}V = -\int_V C_v \frac{\alpha K_T}{\rho C_v}\alpha T \frac{\partial T}{\partial t}\rho \mathrm{d}V$$
$$= -\int_V C_v \gamma \alpha T \frac{\partial T}{\partial t}\rho \mathrm{d}V \qquad [24]$$

where we have used the Grüneisen parameter $\gamma = (\alpha K_T/\rho C_v)$. Using the fact that C_v is approximately constant (the Dulong–Petit–Debye limit for solids at high temperature), we finally obtain

$$E_a = -MC_v\langle\gamma\alpha T\rangle\frac{\mathrm{d}\langle T\rangle}{\mathrm{d}t} \qquad [25]$$

where $\mathrm{d}\langle T\rangle/\mathrm{d}t$ is the average cooling rate. Using $\gamma \approx 1.5$ and $\langle\alpha\rangle T \approx 0.03$, we find that $E_a \approx 1$ TW for a global cooling rate of 100 K Gy^{-1}. As will be shown in the succeeding text, this estimate of the cooling rate must be considered as an upper bound. The adiabatic heating term is smaller than the uncertainty on the energy budget and may be neglected, which leads to

$$MC_v\frac{\mathrm{d}\langle T\rangle}{\mathrm{d}t} = -\int_A \overline{q}_r \mathrm{d}A + \int_V H\mathrm{d}V \qquad [26]$$

which is the secular cooling equation. The average cooling rate lumps together the contributions of two reservoirs with very different physical properties, the mantle and the core, and must be handled with care. For this reason, the two associated energy budgets will be discussed separately.

7.06.2.4 Summary

The thermodynamics of the cooling Earth involve both slow contraction and convective motions. The change of gravitational energy due to thermal contraction does not enter in the global heat budget because it is stored as strain energy. Similarly, the viscous dissipation can be important but is balanced by the bulk buoyancy flux. The secular cooling is dominated by the imbalance between radiogenic heat production and heat loss through the Earth's surface.

7.06.3 Heat Loss Through the Seafloor

For the purposes of calculating the rate at which the Earth is losing heat, the most direct and unbiased method is to integrate individual measurements of heat flux over the surface. As we shall see, this method fails in the oceans, and one has to use theory in order to obtain a reasonable estimate (Jaupart and Mareschal, Chapter 6.05). This approach has been criticized on the grounds that it is a theoretical one and leads to a biased result. We shall discuss this point in detail. Our present understanding of the global thermal budget of the Earth can be summarized in the map of the surface heat flux (Figure 1). It is important to remember that this map is based on the heat flux observations on the continents and their margins but that for the oceans, the surface heat flux is calculated from our best model of energy loss by the oceanic lithosphere.

7.06.3.1 Oceanic Heat Flow Measurements

Heat flow through permeable rock and sediment involves two mechanisms: conduction through the solid matrix and water flow through pores and fractures into the sea. Measuring the latter directly is very costly and requires continuous recording over long timescales in order to determine a representative flow rate. The vast majority of marine heat flux determinations rely on the probe technique, such that a rigid rod carrying a thermistor chain is shoved into sediments. Determination of the temperature gradient and of the sediment thermal conductivity leads to the heat lost by conduction. Another technique is to measure temperatures in deep-sea drillholes. This is clearly the best technique because it relies on measurements over a

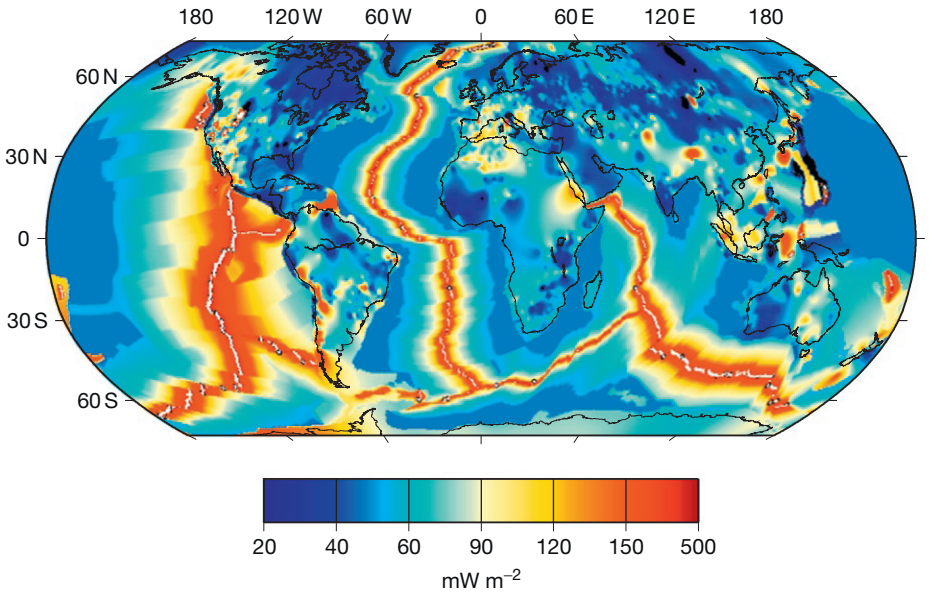


Figure 1 Global heat flux of the Earth combining heat flux measurements on land and continental margins with a thermal model for the cooling of the oceanic lithosphere.

large depth range through poorly permeable crystalline basement, but it is particularly time-consuming. Drilling operations perturb the thermal environment greatly, implying that measurements cannot be done just after drilling is completed and require hole reentry. In addition, the number of deep-sea drillholes is too small to provide a good sampling of the seafloor. The few comparisons that have been made between the two techniques show that the shallow probe technique provides reliable results (Erickson et al., 1975). Obviously, these techniques only account for conductive heat transport, a systematic bias that has important consequences.

The extent of alteration in ophiolitic massifs shows that hydrothermal circulation is pervasive and affects large volumes of oceanic rocks (Davis and Elderfield, 2004). In situ quantitative assessment of heat transport by hydrothermal circulation can only be achieved by measuring the discharge rates and temperature anomalies of hot aqueous upwellings out of the seafloor (Ramondec et al., 2006). A single black smoker can evacuate as much as 0.1 MW (Barreyre et al., 2012), and diffuse venting through thin crack networks and bacterial mats accounts for an even larger power (Mittelstaedt et al., 2012). About 20 MW is lost to the sea through an $\approx 50 \times 50$ m region of the Lucky Strike hydrothermal field of the mid-Atlantic ridge (Mittelstaedt et al., 2012). This shows that, in hydrothermal systems, large amounts of heat loss escape conventional measurement techniques. In such systems, the vagaries of crack sealing and porosity change imply large variations of the shallow thermal structure in both time and space. It therefore does not come as a surprise that oceanic heat flux data exhibit enormous scatter (**Figure 2**). This has presented geophysicists with a major difficulty, all the more as the scatter is largest on young seafloor where the largest heat flow values are recorded.

Various methods have been used to deal with the scatter of heat flow data and the influence of hydrothermal circulation. One has been to bin data by age group in the hope that measurement errors cancel each other in a large data set, but this is not valid if measurement errors are not random, which is the case here. In the widely used compilation by Stein and Stein (1992), heat flux data have been binned in 2 My age intervals (**Figure 2**). This global data set includes a very large number of measurements with no quality assessment. Most of the early data are associated with larger errors than those of recent surveys due to small probe lengths and because thermal conductivity determinations were not made in situ. The first age bin presents a specific problem because it is characterized by the largest heat flux values as well as the most conspicuous signs of hydrothermal activity. A proper average for this age bin requires data at very young ages, <1 My, say, which are virtually nonexistent. Another strategy to measure the oceanic heat loss has been to seek sites with thick sedimentary cover. In crystalline basement, hydrothermal convection proceeds through cracks and generates a highly heterogeneous temperature field with narrow discrete anomalies. With a thick sedimentary cover, upwellings are slowed down and become diffuse and hence tend toward thermal equilibrium with the surrounding matrix. In this case, conduction is the dominant heat transport mechanism and conventional heat flow techniques yield reliable results. Heat flux varies spatially but the integrated value is equal to the heat extracted from the lithosphere. A reliable heat flux determination therefore requires closely spaced stations over a representative area. A third strategy has been to enhance our understanding of heat transport characteristics through young seafloor, which was done with both detailed local surveys and physical models (Davis et al., 1999; Spinelli and Harris, 2011).

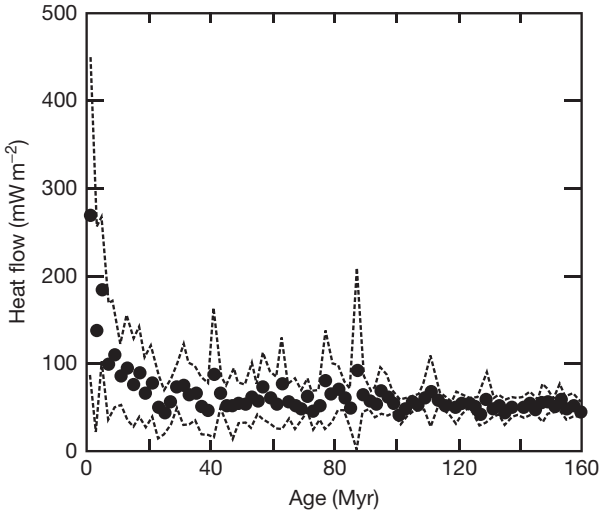


Figure 2 Distribution of heat flux data as a function of age from the compilation by Stein and Stein (1992). Dots represent averaged heat flux values in 2 My bins. Dashed lines indicate the envelope at one standard deviation.

In summary, oceanic heat flux data are plagued by systematic errors that arise from the measurement environment and that can only be sorted out by careful small-scale local studies. There are very sound reasons that explain why heat flux data underestimate the total heat flux out of the seafloor, especially at young ages (Harris and Chapman, 2004). Using the raw data average turns a blind eye to this fundamental problem and pays no attention to the systematics of the scatter. For the purposes of calculating heat loss through the oceans, such a brute force method is not appropriate and one must instead account carefully for the specifics of the shallow oceanic environment. We shall rely on detailed heat flux surveys in selected areas and shall generalize the insights gained in this manner by comparing the data to robust thermal models for the cooling of oceanic lithosphere. We shall also evaluate the amplitude of the scatter in heat flow data using simple thermal considerations. We shall finally show how the topography of the seafloor records the amount of heat that is lost by the oceanic lithosphere.

7.06.3.2 Cooling of the Oceanic Lithosphere

Here, we are interested in the large-scale thermal structure and evolution of oceanic lithosphere and momentarily ignore the complexities of the shallow environment. The heat equation provides the basic framework

$$\rho C_p \frac{DT}{Dt} = \rho C_p \left(\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T \right) = \nabla \cdot (\lambda \nabla T) \qquad [27]$$

where C_p is the heat capacity, ρ is the density of the lithosphere, λ is thermal conductivity, and $\mathbf{v}$ is the velocity of the plate. We have neglected radiogenic heat production, which is very small in

the oceanic crust and in mantle rocks (see succeeding text), and viscous heat dissipation. Over the temperature range of interest here, variations of heat capacity amount to ±20%. Such subtleties will be neglected momentarily for the sake of clarity and simplicity. They must be taken into account, however, for accurate calculations (McKenzie et al., 2005). In the upper boundary layer of a convection cell (see **Chapter 7.04**), vertical advective heat transport is negligible. Over the large horizontal distances involved, vertical temperature gradients are much larger than horizontal ones, except in the immediate vicinity of the ridge axis, and one may neglect horizontal diffusion of heat. The validity of this standard boundary layer approximation was verified by Sclater et al. (1980), who found that calculations with and without horizontal heat diffusion could not be distinguished from one another at ages larger than 1 My.

It is best to work in a reference frame that moves with the plate (i.e., a Lagrangian approach). Over the lifetime of an oceanic plate, secular cooling may be safely neglected so that temperature can be written as a function of age, noted τ, instead of time. For negligible horizontal heat transport, the heat eqn [27] can be reduced to

$$\frac{\partial T}{\partial \tau} = \kappa \frac{\partial^2 T}{\partial z^2} \qquad [28]$$

where κ is thermal diffusivity. This is the one-dimensional heat diffusion equation, whose solution requires a set of initial and boundary conditions. The initial condition requires specification of the thermal structure of an oceanic spreading center. For the top boundary condition, one must in principle account for thermal conditions in the shallow oceanic environment. The bottom boundary condition depends on the efficacy of heat transport in the asthenosphere. All these conditions are only approximations to more complex realities and must be evaluated carefully.

7.06.3.2.1 The shallow oceanic environment

The upper part of young oceanic lithosphere is made of fractured crust allowing pervasive water circulation, which is not accounted for in the heat eqn [27]. An exact solution must deal with different environments and involve a host of parameters that are not known precisely and vary with time, such as permeability, which decreases as fractures and pores get clogged by precipitates from the highly charged solutions that circulate through them. Fortunately, several facts conspire to dramatically reduce the impact of water circulation on bulk lithospheric cooling.

Hydrothermal convection in the oceanic crust is self-defeating and becomes inefficient after about 0.1 My (Dunn et al., 2000; McLennan et al., 2005; Spinelli and Harris, 2011). The end result is cold oceanic crust sitting on basement that has not had time to lose much heat by conduction. After this initial transient, the cold crust heats up against the hotter basement by diffusion and temperatures in both crust and basement adjust to a profile resembling that of steady conductive cooling from $\tau = 0$. On the ridge flanks, away from the axial hydrothermal zone, the highly permeable pillow lava layer at the top of the crust hosts strong lateral water flow (Davis and Elderfield, 2004) that effectively smoothes out thermal anomalies and imposes a nearly constant and uniform temperature. Below this aquifer, heat conduction is the dominant process, save for isolated deep open fractures. Fluid-driven cooling therefore proceeds in two different phases, with deep convection in large parts of the crust followed by superficial circulation in a thin aquifer. Only the first phase is significant for the large-scale heat budget of the lithosphere, but it has a prolonged influence on the shallow lithospheric thermal structure due to the slowness of diffusive relaxation. It may leave its imprint on the surface heat flux for as long as 10 My depending on the thickness affected by axial hydrothermal convection (Spinelli and Harris, 2011).

What has been achieved by hydrothermal convection at the ridge axis is rapid heat extraction out of the crust at the top of the lithosphere, and this could be taken as an initial condition for the conduction cooling model. The vagaries of convection, fracturing, and sealing in the crust and of sedimentation in a ridge environment with irregular topography prevent a universally valid axial temperature structure, however, and the preferred approach has been to use a 'reference' conduction model with simple boundary and initial conditions. We shall see that this allows an excellent fit to various types of data. Departures from model predictions can be calculated, observed, and included in an error analysis.

Due to the high efficacy of heat transport in seawater, the seafloor is at a fixed temperature of about 4 °C (in practice 0 °C for convenience), which provides an upper boundary condition.

7.06.3.2.2 Initial condition: Temperature distribution at the ridge axis

To specify the temperature profile at the ridge axis, which provides the initial condition for the cooling model, two effects come into play. Mantle that rises toward the oceanic ridge undergoes pressure release and partial melting, which absorbs latent heat. This mantle is hotter than its surroundings and loses heat laterally by diffusion. During isentropic ascent of dry mantle, temperature decreases by about 200 K (McKenzie and Bickle, 1988), which is small relative to the temperature contrast at the top of the oceanic plate ($\approx$1350 K). Thus, it is commonly assumed that the axial temperature does not vary with depth and is equal to a constant value noted T_M. Calculations based on a realistic temperature profile lead to minor differences (McKenzie et al., 2005). One problem with these temperature profiles is that they specify a finite temperature drop at the surface at $\tau = 0$, implying that the surface heat flux diverges as $\tau \to 0$. This problem disappears if one specifies that the heat that is advected vertically by the upwelling mantle at the ridge is lost by horizontal conduction to the adjacent oceanic plate (Davis and Lister, 1974). In this modified model, the heat flux is finite at the ridge but differs significantly from the previous model only for ages < 1 My (Sclater et al., 1980). For determining the heat loss, such differences are of no consequence: heat flux integrated over finite areas remains bounded regardless of the initial condition and is very weakly sensitive to the choice of initial condition (Sclater et al., 1980).

7.06.3.2.3 Bottom boundary condition

The simplest model, called the 'half-space' model, has the lower boundary at infinite depth and assumes that temperature remains finite, such that cooling proceeds unhampered over the entire age span of oceanic lithosphere. For the reference

upper boundary condition, such that $T(z=0)=0$, the temperature distribution is then

$$T(z,\tau)=T_{\mathrm{M}}\,\mathrm{erf}\left(\frac{z}{2\sqrt{\kappa\tau}}\right)=\frac{2T_{\mathrm{M}}}{\sqrt{\pi}}\int_0^{z/2\sqrt{\kappa\tau}}\exp\left(-\eta^2\right)\mathrm{d}\eta \qquad [29]$$

for which the surface heat flux is

$$q(0,\tau)=\frac{\lambda T_{\mathrm{M}}}{\sqrt{\pi\kappa\tau}}=C_{\mathrm{Q}}\tau^{-1/2} \qquad [30]$$

where C_{Q} is a constant. This equation makes the very simple prediction that heat flux varies as $\tau^{-1/2}$. One remarkable feature is that this relationship holds for arbitrary temperature-dependent physical properties (Carslaw and Jaeger, 1959; Lister, 1977) (**Appendix D**). Numerical models of mantle convection that are in a platelike regime conform very well to this relationship. **Figure 3** displays a snapshot of the temperature field and surface observables in such a model, from Grigné et al. (2005). One sees clearly that the horizontal velocity at the surface is piecewise constant, defining plates, and that the heat flux decreases with distance x from ridges as $1/\sqrt{x}$, that is to say $\tau^{-1/2}$ for a constant velocity. We shall see that the value for the mantle temperature T_{M} remains subject to some uncertainty. The value of the constant C_{Q} in eqn [30], however, may be determined empirically from the data, as will be shown later.

In the other class of models, a boundary condition is applied at some depth, which marks the base of the 'plate.' In principle, one should solve for heat supply from the asthenosphere. However, this requires elaborate physical models of mantle convection relying on specific choices of material properties and simplifying assumptions. For the sake of simplicity, one may consider two simple end-member cases, such that temperature or heat flux is constant at the base of the plate. Both these boundary conditions are approximations. For example, the fixed temperature boundary condition requires infinite thermal efficiency for heat exchange between the plate and the mantle.

The choice of the proper basal boundary condition is important because it determines the relationship between the relaxation time and the plate thickness. For plate thickness a, the diffusion relaxation time scales with a^2/κ with a proportionality constant that depends on the bottom boundary condition. Specifically, the thermal relaxation time of a plate is four times as long for fixed heat flux at the base than it is for fixed temperature. Thus, for the same relaxation time, a plate with a fixed basal temperature is twice as thick as one with a fixed basal heat flux. For short times, the cooling rate (i.e., the surface heat flux) does not depend on the lower boundary condition and, for both types of plates, it is the same as the heat flux for the cooling half-space. The details are provided in **Appendix E**. Thus, it is better to use a half-space model for young ages because it relies on a reduced set of hypotheses and because it does fit the oceanic data, as will be shown later. Furthermore, it has been tested over and over again and forms the basis for scaling laws of convective heat flux in many different configurations (Howard, 1964; Olson, 1987; Turcotte and Oxburgh, 1967). This simple model breaks down at ages larger than about 80 My for reasons that are still debated. For this reason, it may be wise not to rely on a specific physical model, such as that of a plate. For ages >80 My, we do not need a theoretical cooling model to calculate the oceanic heat loss because the data scatter is small. For this second phase of lithospheric evolution, the heat flux is approximately constant (see succeeding text), indicating that heat is supplied to the base of the oceanic lithosphere and that thermal steady state is nearly achieved.

One final worry is that oceanic ridges and oceanic plates drift over mantle that may not be thermally well mixed, which could be responsible for a bottom boundary condition that varies with age. A detailed study of oceanic tholeiitic basalts, however, demonstrates that their composition, and hence the temperature of the mantle from which they were derived, has not changed significantly for about 80 My (Humler et al., 1999).

7.06.3.3 Validating the Cooling Model for Oceanic Lithosphere

7.06.3.3.1 High-resolution surveys near mid-ocean ridges

A very detailed heat flux survey on young seafloor near the Juan de Fuca Ridge was conducted by Davis et al. (1999) with three specific goals: to evaluate the intensity and characteristics of hydrothermal circulation, assess local thermal perturbations due to basement irregular topography, and test cooling models for the lithosphere. **Figure 4** shows the salient results. Direct observation of the seafloor shows that the basement outcrop region is a zone of recharge and that, beneath the sedimentary cover, water flow is dominantly horizontal. Focused discharge occurs at a few locations in association with basement

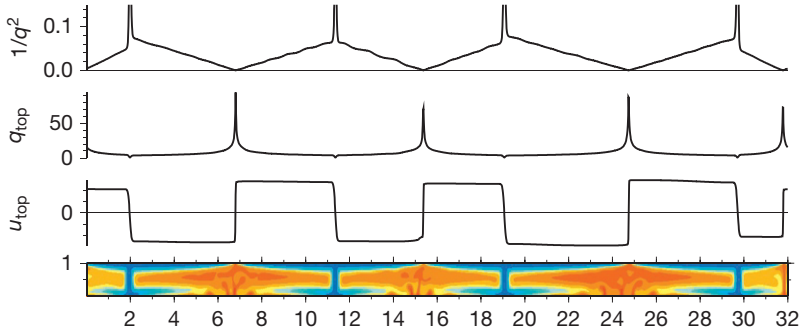


Figure 3 Snapshot of temperature, surface velocity (u_{top}), heat flux (q), and pseudo age ($1/q^2$) in a numerical convection model with self-consistent plate tectonics (Grigné et al., 2005). See **Appendix H** for details. Note that the pseudo age varies linearly as function of distance to the ridge, which is consistent with the $\tau^{-1/2}$ heat flux law for young oceanic lithosphere.

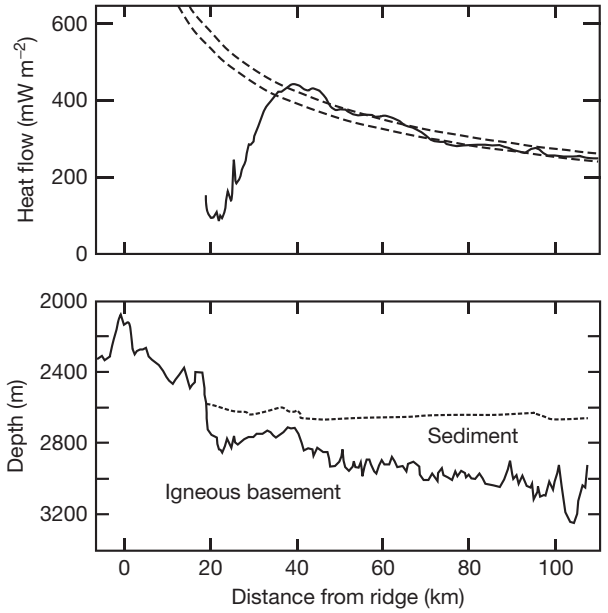


Figure 4 High-resolution heat flux profile near the Juan de Fuca Ridge from Davis et al. (1999). Dashed lines stand for two predictions of the half-space cooling model with constant C_Q in the $\tau^{-1/2}$ heat flux–age relationship equal to 470 and 510 (with heat flux in mW m^{-2} and age in My).

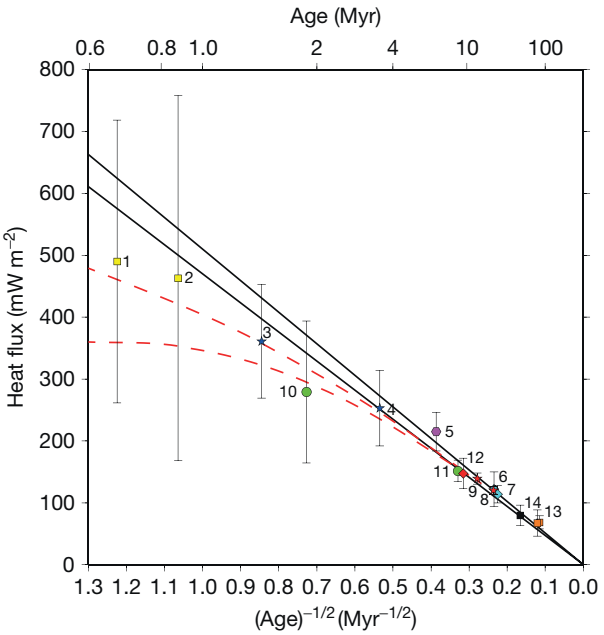


Figure 5 Heat flow versus age for selected sites with a sediment blanket. 1–2: Galápagos Spreading Center (Williams et al., 1974); 3–4: Juan de Fuca Ridge (Hasterok et al., 2011); 5: Costa Rica Rift (Hasterok et al., 2011); 6–7: Cocos Plate (Hasterok et al., 2011); 8–9: the Gulf of Aden (Cochran, 1981; Lucazeau et al., 2010); 10–11 East Pacific Rise (Villinger et al., 2002); 12: south Azores (Lucazeau et al., 2006); 13: Mascarene basin (Bonneville et al., 1997); 14: Reykjanes Ridge (Sclater and Crowe, 1979). Several sites (1, 2, 3, and 10) deviate from the conductive cooling trend, which can be explained by thermal relaxation of the initial thermal structure at the ridge inherited from pervasive hydrothermal convection in the crust. Dashed red curves represent the predictions of models of hydrothermal convection in fast spreading centers extending to depths of 2 and 6 km (Spinelli and Harris, 2011).

topographic highs. Elsewhere, water flow is diffuse, such that it equilibrates with the sediments.

At the Juan de Fuca Ridge site, the heat flux fluctuates on the scale of station spacing (2 km), which implies that surveys with coarser spatial resolution do not yield meaningful results. To emphasize the long-wavelength trends, a 15 km running average was applied on the heat flux profile. This profile shows two distinct trends: near the basement outcrop, to the left, the heat flux increases with age and reaches a maximum value in excess of 400 mW m^{-2}, and further away from the outcrop, the heat flux decreases with age. In the latter region, the total heat flux variation is large enough to allow comparison with theoretical models for the cooling of oceanic lithosphere. The data conform to the $\tau^{-1/2}$ relationship and indicate that constant C_Q in eqn [30] is between 470 and 510 (with heat flux in mW m^{-2} and age in My).

7.06.3.3.2 Global analysis

The local Juan de Fuca Ridge survey can be corroborated independently using filtered heat flow data over larger age ranges. Hasterok et al. (2011) used a revised heat flow database with more than 15 000 oceanic values and selected sites located >60 km from the nearest seamount where the sediment cover is >400 m. These data exhibit a well-defined correlation with the age of the seafloor but remain scattered.

For better results, we focus on 14 sites with thick sedimentary cover and dense heat flow sampling (**Figure 5**). For those sites, hydrothermal circulation that may still be active in the igneous basement is effectively sealed off by the hydraulically resistive sedimentary cover. There are no localized discharge zones and the average heat flux is equal to the rate at which the basement loses energy. The data conform to the $\tau^{-1/2}$ relationship between 3 and 80 My. In order to determine the best-fit coefficient, we add the constraint that the heat flux tends to zero as age tends to infinity and find that $C_Q = 490$, close to the value for the Juan de Fuca Ridge. For ages <3 My, the measurements are more dispersed and site-averaged values fall below the simple cooling model. This can be explained by the conductive relaxation of the anomalous thermal structure generated by hydrothermal convection in the axial zone. We show in **Figure 5** calculated values from the model of Spinelli and Harris (2011), which account nicely for the data.

Combining the local Juan de Fuca Ridge survey and the global data set of reliable heat flux sites, we conclude that $C_Q = 490 \pm 20$, with an uncertainty of ±4%. **Table 3** compares the various estimates that have been used in the past. The heat loss estimate of Pollack et al. (1993) was based on $C_Q = 510$, which is clearly an upper bound. This value was taken from the analysis of Stein and Stein (1992), which was based on the plate model with constant basal temperature. One feature of this model is that $T_M = 1725$ K, a high value that is not consistent with the average ridge axis temperature derived from the compositions of mid-ocean ridge basalts (MORBs) (Kinzler and Grove, 1992) (**Table 4**).

The $\tau^{-1/2}$ relationship is based on sound physics and holds for conductive cooling models at early times, including those

Table 3 Estimates of the continental and oceanic heat flux and global heat loss

	Continental (mW m^{-2})	*Oceanic (mW m^{-2})*	*Total (TW)*
Williams and von Herzen (1974)	61	93	43
Davies (1980a,b)	55	95	41
Sclater et al. (1980)	57	99	42
Pollack et al. (1993)	65	101	44
Davies and Davies (2010)	71	105	47
This study[a]	65	94	46

[a]The average oceanic heat flux does not include the contribution of hot spots. The total heat loss estimate does include 3 TW from oceanic hot spots.

Table 4 Potential temperature of the oceanic upper mantle

	Reference	*Method*
1333 °C[a]	Parsons and Sclater (1977)	Average depth + heat flux
1450 °C[a]	Stein and Stein (1992)	Average depth + heat flux
1300–1370 °C[a]	Carlson and Johnson (1994)	True basement depth (DSDP)
1315 °C[a]	McKenzie et al. (2005)	Depth + heat flux with λ(T), C_p(T) and α(T)
1280 °C[a]	McKenzie and Bickle (1988)	Average basalt composition
1315–1475 °C[a]	Kinzler and Grove (1992)	Basalt composition
1275–1375 °C[a]	Katsura et al. (2004)	Isentropic profile through the Ol–Wa phase change
1454 ± 81 °C	Putirka et al. (2007)	Olivine–liquid equilibriums

[a]Temperature estimate for a cooling model with constant temperature below the ridge axis (i.e., which does not account for isentropic decompression melting).

for a plate of fixed thickness. This relationship provides a good fit to the data and we shall use it for the 3–80 My age range. For ages <3 My, this relationship breaks down for the global data set but is still valid in some areas (**Figure 4**). We shall still use it and calculate the associated uncertainty. For ages >80 My, we shall simply use a constant heat flux as the data indicate that $q_{80} \approx 48$ mW m^{-2} (Lister et al., 1990) (**Figure 6**).

7.06.3.3.3 Depth of the seafloor

The variation of ocean floor depth with age provides an additional constraint on lithospheric cooling at the cost of introducing a new physical property, the coefficient of thermal expansion. As for heat flow, depth data are highly scattered due to intrinsic seafloor roughness, the presence of seamounts, and a variable sediment cover. Coping with the noisy data set has involved again either binning of the data (Parsons and Sclater, 1977) or rejecting 'anomalous' sites (Crosby and McKenzie, 2009; Crough, 1983; Hillier and Watts, 2005; Korenaga and Korenaga, 2008). Analyzing seafloor subsidence in detail would be outside the scope of this chapter and we only seek confirmation for the two main features of lithospheric evolution that are used here: the $\tau^{-1/2}$ heat

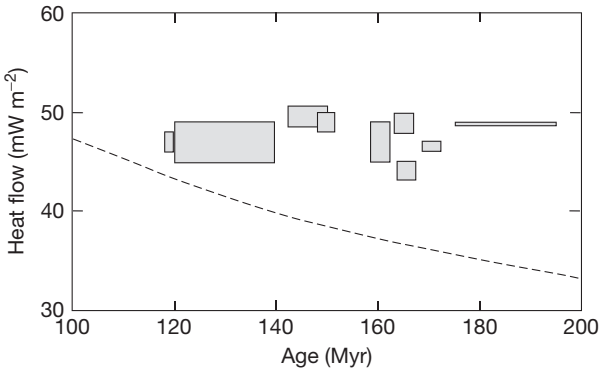


Figure 6 Heat flux data and prediction of the half-space cooling model for ages larger than 100 My. Reproduced from Lister CRB, Sclater JG, Nagihara S, Davis EE, and Villinger H (1990) Heat flow maintained in ocean basins of great age – Investigations in the north-equatorial West Pacific. *Geophysical Journal International* 102: 603–630.

flux–age dependence up to 80 My and the interrupted cooling at later ages.

An isostatic balance condition leads to a very simple equation for subsidence with respect to the ridge axis (Sclater and Francheteau, 1970):

$$\Delta h(\tau) = h(\tau) - h(0) = \frac{1}{\rho_m - \rho_w} \int_0^d (\rho[T(z,\tau)] - \rho[T(z,0)])\mathrm{d}z \qquad [31]$$

where $h(\tau)$ and $\Delta h(\tau)$ are the depth of the ocean floor and subsidence at age τ and where ρ_m and ρ_w denote the densities of mantle rocks at temperature T_M and water, respectively. In this equation, d is some reference depth in the mantle below the thermal boundary layer. This equation neglects the vertical normal stress at depth d, which may be significant only above the mantle upwelling structure, that is, near the ridge axis. We are interested in the heat flux out of the seafloor, $q(0,\tau)$. Assuming for simplicity that the coefficient of thermal expansion α is constant, the equation of state for near-surface conditions is

$$\rho(T) = \rho_m[1 - \alpha(T - T_M)] \qquad [32]$$

From the isostatic balance equation [31], we obtain

$$\frac{\mathrm{d}h}{\mathrm{d}\tau} = \frac{-\alpha\rho_m}{\rho_m - \rho_w} \frac{\mathrm{d}}{\mathrm{d}\tau}\left[\int_0^d T(z,\tau)\mathrm{d}z\right] = \frac{-\alpha}{C_p(\rho_m - \rho_w)} \frac{\mathrm{d}}{\mathrm{d}\tau}\left[\int_0^d \rho_m C_p T(z,\tau)\mathrm{d}z\right] \qquad [33]$$

where we have also assumed that C_p is constant. Heat balance over a vertical column of mantle between $z=0$ and $z=d$ implies that

$$\frac{\mathrm{d}h}{\mathrm{d}\tau} = \frac{\alpha}{C_p(\rho_m - \rho_w)}[q(0,\tau) - q(d,\tau)] \qquad [34]$$

which states that thermal contraction reflects the net heat loss between the surface and depth d. Because $q(0,\tau)$ depends on T_M, the subsidence rate also depends on the initial temperature at the ridge axis. This equation states that the surface heat flux is the sum of heat flux at depth d and the amount of cooling over vertical extent d. Using only the latter therefore leads to an underestimate of the surface heat flux.

Carlson and Johnson (1994) investigated these theoretical predictions using the best data set, basement depths from deep-sea drillholes, which require no correction for sediment thickness. The data are consistent with the half-space cooling model up to an age of about 80 My and indicate that the rate of subsidence decreases markedly at later ages. In its simplest formulation with constant physical properties and a constant temperature beneath the ridge, the plate model does not allow a good fit to the data. Using the best-fit parameters deduced from subsidence data and estimates for the various physical properties of mantle rocks in eqn [34] (i.e., for α, C_p, and the densities), Carlson and Johnson (1994) predicted heat flux values that were consistent with reliable heat flux data. For ages <80 My, the best-fit depth versus age relationship is

$$h(\tau) = (2600 \pm 20) + (345 \pm 3)\tau^{1/2}, \text{with } h \text{ in meters and } \tau \text{ in My} \quad [35]$$

From this relationship, the predicted heat flux over the same age range is

$$q(0, \tau) = (480 \pm 4)\tau^{-1/2}, \text{with } q \text{ in mW m}^{-2} \text{ and } \tau \text{ in My} \quad [36]$$

We shall return to these estimates later. Using the same physical properties, Carlson and Johnson (1994) found that fitting the depth data with the plate model and half-space model leads to different values for the average ridge axis temperature T_M: 1470 and 1370 °C, respectively.

Many authors have debated on how 'normal' seafloor subsides using various filtering techniques and by winnowing data from seamounts and other anomalous areas. The major issue was whether or not the data truly indicate flattening of the seafloor for ages larger than 80–100 My (Crough, 1983; Korenaga and Korenaga, 2008). Recent work has established that flattening does occur (Crosby and McKenzie, 2009; Hillier, 2010; Hillier and Watts, 2005). For our present purposes, this is important because it confirms what is indicated by heat flow data: after about 80 My, cooling of the oceanic lithosphere slows down markedly, and thermal steady state gets established due to heat supply from the asthenosphere.

7.06.3.4 Heat Loss Through the Ocean Floor

Simple physical considerations demonstrate that seafloor topography and surface heat flux record the same phenomenon and furthermore that the subsidence rate is proportional to surface heat flux for as long as 80 My. In contrast, the raw oceanic heat flux data set records the conductive heat flux through a heterogeneous permeable superficial sedimentary layer. Using this data set leaves the seafloor topography unexplained.

The plate model with fixed basal temperature has been used to calculate the total oceanic heat loss by several authors (Hasterok, 2013; Pollack et al., 1993; Sclater et al., 1980; Stein and Stein, 1992). Its advantage is that it provides a function that fits noisy data over the whole range of ages. We shall not follow this approach for two reasons. One is that the plate model is only an approximation of the physics involved. Maintaining a constant temperature at a fixed depth requires infinitely efficient heat exchange between lithosphere and asthenosphere and a mechanism that maintains the plate thickness. Another reason is that the plate model depends on parameters such as the plate thickness and axial ridge temperature that are determined a posteriori by a best fit to the data. This procedure may introduce trade-offs such that, for example, data at old ages steer model predictions away from data at younger ages. We prefer to stay as close to the data as possible with as few hypotheses as possible. To calculate the total rate of oceanic heat loss, we separate the data in two different age intervals: 0–80 My, where the $\tau^{-1/2}$ law holds, and older seafloor where heat flux data depart from the half-space model. For seafloor older than 80 My, the heat flux is approximately constant $q_{80} \approx 48\,\text{mW m}^{-2}$ (Lister et al., 1990) (**Figure 6**). Deviations from this value are $\leq 3\,\text{mW m}^{-2}$ and exhibit no systematic age trend. The mean is determined with an uncertainty of $1\,\text{mW m}^{-2}$, which represents 1% of the average oceanic heat flux. This has negligible impact on the total heat loss estimate, which is dominated by the young seafloor contribution. We also use bathymetry data for seafloor younger than 80 My but not for older ages. We add the contribution of marginal basins that follow the standard oceanic heat flux model, as demonstrated by Sclater et al. (1980). Continental margins, passive or active, are transition zones between continents and oceans and are treated in the continental section.

7.06.3.4.1 Global oceanic heat loss calculation

Heat loss through the ocean floor is equal to

$$Q_0 = \int_0^{\tau_{max}} q(0, \tau)\frac{dA}{d\tau}d\tau \quad [37]$$

where $A(\tau)$ is the distribution of seafloor with age, which can be deduced from maps of the ocean floor (Müller et al., 1997, 2008; Royer et al., 1992; Sclater et al., 1980). Using the data sets of Royer et al. (1992) and Müller et al. (1997), Rowley (2002) and Cogné and Humler (2004) found that a simple linear relationship provides a good fit to the data (**Figure 7**), confirming the earlier result of Sclater et al. (1980):

$$\frac{dA_1}{d\tau} = C_A(1 - \tau/\tau_m) \quad [38]$$

These three different groups of authors agree that $\tau_m = 180$ My but quote slightly different values of the coefficient C_A: $3.45\,\text{km}^2\,\text{year}^{-1}$ for Sclater et al. (1980), $2.96\,\text{km}^2\,\text{year}^{-1}$ for Rowley (2002), and $2.85\,\text{km}^2\,\text{year}^{-1}$ for Cogné and Humler (2004). The high estimate of Sclater et al. (1980) is due to the inclusion of marginal basins, which contribute $\approx 0.38\,\text{km}^2\,\text{year}^{-1}$ to the global accretion rate. The small difference of about $0.12\,\text{km}^2\,\text{year}^{-1}$ (4%) between the more recent estimates of Rowley (2002) and Cogné and Humler (2004) arises from the different methods used to fit the data. One independent constraint is brought by the total area of ocean floor, which is sensitive to the exact location of the continent–ocean boundary. A detailed analysis of continental margins leads to a total continental area of $210 \times 10^6\,\text{km}^2$ (Cogley, 1984) (see **Table 5**). From this, the total seafloor surface is $300 \times 10^6\,\text{km}^2$, slightly less than the value used by Sclater et al. (1980). For a triangular age distribution with a maximum age $\tau_m = 180$ My, this implies that $C_A = 3.34\,\text{km}^2\,\text{year}^{-1}$. Subtracting the contribution of marginal basins, this corresponds exactly to the Rowley (2002) estimate. Thus, for our purposes, we shall use $C_A = 3.34\,\text{km}^2\,\text{year}^{-1}$. This discussion illustrates that uncertainties may come from unexpected variables, the area of the

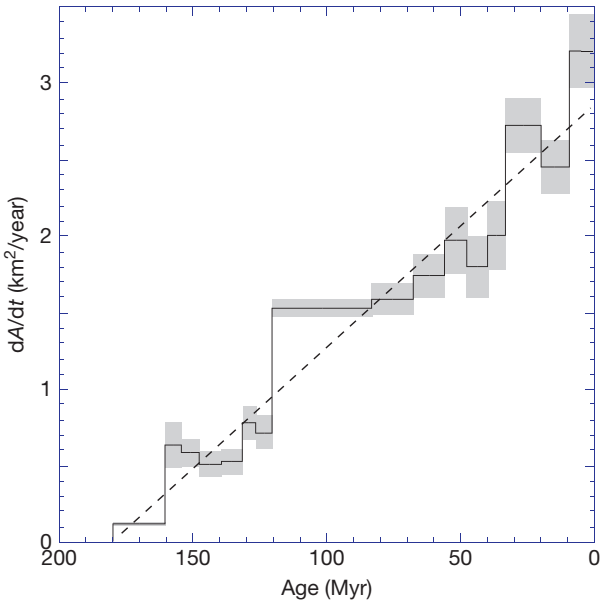


Figure 7 Distribution of seafloor ages. The dashed line is the best-fit linear function. Modified from Cogné J-P and Humler E (2004) Temporal variation of oceanic spreading and crustal production rates during the last 180 My. *Earth and Planetary Science Letters* 227: 427–439.

Table 5 Surface area and heat flux in oceans and continents

	Area	*Total heat flux*
Oceans		
Oceanic	273×10^6km^2	
Marginal basins	27×10^6km^2	
Total oceans	300×10^6km^2	32 TW
Continents		
Precambrian	95×10^6km^2	
Paleozoic	23×10^6km^2	
Stable continents	118×10^6km^2	
Active continental	30×10^6km^2	
Submerged (margins and basins)	62×10^6km^2	
Total continental	210×10^6km^2	14 TW

seafloor in this particular instance. The continental heat flux budget must account for the remaining 210×10^6 km^2 (Table 5).

Integrating separately seafloor younger and older than 80 My gives

$$Q_{80-}=\int_0^{80} C_Q\tau^{-1/2}C_A(1-\tau/180)\mathrm{d}\tau=24.3\mathrm{TW} \quad [39]$$

$$Q_{80+}=q_{80}\int_{80}^{180} C_A(1-\tau/180)\mathrm{d}t'=4.4\mathrm{TW} \quad [40]$$

$$Q_{\mathrm{oceans}}=29\pm1\mathrm{TW} \quad [41]$$

where the uncertainty comes mostly from that on coefficient C_Q. The present estimate is slightly less than earlier estimates because of the slightly lower ridge temperature (or equivalently, the slightly smaller value of coefficient C_Q in the heat flux vs. age relationship) and because of the revised estimate for the mean accretion rate at zero age C_A. For $C_Q=510\,\mathrm{mW\,m^{-2}\,My^{-1/2}}$ and $C_A=3.45\,\mathrm{km^2\,year^{-1}}$, the heat loss would be 31 TW. In an independent study, Hasterok (2013) relied on a 'heat flow calibrated plate model' and obtained 29.4 TW, which is identical to our result.

For young ages, the bathymetry provides a direct measure of the heat lost by the cooling plate. We obtain another heat loss estimate with the following equation:

$$Q(80-)=\frac{C_p}{\alpha}(\rho_m-\rho_w)\int_0^{80}\frac{\mathrm{d}h}{\mathrm{d}\tau}\frac{\mathrm{d}A(\tau)}{\mathrm{d}\tau}\mathrm{d}\tau \quad [42]$$

$$\approx 24\mathrm{TW} \quad [43]$$

which is formally identical to the heat flux equation above and where $\rho_m=3300\,\mathrm{kg\,m^{-3}}$, $C_p=10^3\,\mathrm{J\,kg^{-1}\,K^{-1}}$, and $\alpha=3\times10^{-5}\,\mathrm{K^{-1}}$. In old basins, where the heat flux $\approx48\,\mathrm{mW\,m^{-2}}$, the bathymetry is almost flat and cannot be used to estimate the rate of heat loss.

Uncertainties on the oceanic heat loss come from three sources other than that on the coefficient C_Q. One source is the uncertainty on the age distribution and the total area of ocean floor or, more precisely, the total area of continental shelves. The latter is $<3\%$ and implies an even lower uncertainty on the global heat loss because a change in the area of oceans is compensated by a change in the area of continents. Considering the difference between the average oceanic heat flux and continental heat flux, the resulting uncertainty on the global heat loss estimate is only 1%. The impact of departures from the triangular age distribution is best assessed by comparing the heat loss estimate derived from eqn [37] and that obtained by adding the individual contributions of each age group. Parsons (1982) showed that this difference amounts to about 0.3% of the total, which may be considered negligible. As we shall see later on, however, evaluating the uncertainty on the age distribution must be done over a large timescale and involves consideration of the stability of the mantle convection planform. The two other sources of uncertainty are discussed next and involve the thermal regime of young seafloor, which is not captured accurately by the conductive cooling model, and the contribution of hot spots.

7.06.3.4.2 Heat loss through young ocean floor

The conductive cooling model is vindicated for ages older than 3 My by a global data set of reliable heat flux determinations and on younger seafloor by local surveys with small station spacing and favorable environmental conditions. It is not valid for many areas of the seafloor that are younger than 3 My and particularly for axial hydrothermal zones, however, because it does not account for ridge processes. Yet it is within that age interval that the heat flux is highest, and we evaluate the consequences for heat loss estimates. We cannot determine precisely the error because this would require models for all the different axial hydrothermal zones that exist in the world's oceans. Instead, we compare two different estimates of the energy that is lost to the sea in 3 My through a unit area of seafloor. The estimate from the conductive cooling model is $\approx55\times10^{12}\,\mathrm{J\,m^{-2}}$. For the other estimate, we consider the two independent processes that produce the cooling of young oceanic lithosphere. For hydrothermal convection at the ridge, an upper bound is obtained by considering that the whole crust

gets cooled to the sea temperature. For a 6 km thick crust, this represents $\approx 25 \times 10^{12}\,\mathrm{J\,m^{-2}}$. After this very short-lived phase, conductive cooling is recorded by reliable heat flux measurements on ridge flanks (Figure 5). Over 3 My, these data account for $\approx 40 \times 10^{12}\,\mathrm{J\,m^{-2}}$. Adding these two components leads to $\approx 65 \times 10^{12}\,\mathrm{J\,m^{-2}}$, about 18% higher than the conductive model. This represents the upper bound for the uncertainty. The 0–3 My age range accounts for 18% of the total oceanic heat loss, implying that the final error on the total heat loss estimate is $<3\%$.

7.06.3.4.3 Hot spots

The oceanic heat loss estimate relies on a thermal model that specifies explicitly that the oceanic lithosphere cools by conduction over a well-mixed asthenosphere. This ignores the contribution of isolated hot spots that are the surface expressions of localized mantle upwellings with temperatures that are above those of the well-mixed mantle. Theoretical physical models show that, beneath a moving plate, plume material gets deflected by the ambient mantle flow and spreads at the base of the lithosphere. Thus, the magnitude of the heat flux perturbation at the base of the plate is smaller than that for a vertical plume rising through a static fluid (Olson, 1990; Ribe and Christensen, 1994). The thermal anomaly propagates by diffusion to the top of the plate in a time that depends on the plate thickness and on the extent of plume penetration (i.e., the amount of plate thinning). Given the scatter inherent in oceanic measurements, it is not surprising that heat flux anomalies over hot spot tracks have been elusive. Bonneville et al. (1997) had shown that dense station spacings are required to isolate a meaningful signal from shallow topographic effects and the influence of water circulation. They did find that the Reunion hot spot track was an area of enhanced heat flux, corroborating the earlier findings of Courtney and White (1986) over the Cape Verde Rise. Faced once again with the shortcomings of raw heat flux data, one has had to resort to indirect methods. The heat flux from hot spots has been estimated from the buoyancy of bathymetric swells (Davies, 1988; Sleep, 1990). These estimates are in the range 2–4 TW and are added to the heat loss due to plate cooling.

The hot spot contribution must be handled with care as some mantle plumes may not penetrate the oceanic lithosphere over large vertical distances and may only heat up a thin basal region. In this case, the induced thermal anomaly remains 'locked' within the lithosphere with no surface heat flux expression. The heated region returns with the plate into the mantle in subduction zones and contributes no net heat loss.

7.06.3.5 Summary

Heat loss through the ocean floor cannot be determined using the raw heat flux data set that includes many measurements that are affected by hydrothermal circulation and irregularities of the sediment cover. For ages in the 3–80 My range, predictions of the 'half-space' model for the cooling of the lithosphere can be compared successfully to measurements in selected environments where the effects of hydrothermal circulation can be assessed and accounted for. This model implies values for the mantle temperature beneath mid-ocean ridges that are consistent with independent petrologic models for basalt genesis. Finally, it is consistent with the evolution of seafloor bathymetry that the raw heat flux data leave unaccounted for. The heat loss estimate requires an accurate areal distribution of seafloor ages. Uncertainty in the end result is essentially due to errors on the extent of continental margins. Accounting for the various uncertainties involved, the present-day rate of heat loss through the ocean floor is 32 ± 2 TW. This estimate includes the enhanced heat flux over hot spots.

7.06.4 Heat Loss Through Continents

7.06.4.1 Average Continental Heat flux and Heat Loss Through Continental Areas

There are more than 35 000 heat flux measurements over the continents and their margins, including measurements made in oil wells. Large areas (Antarctica, Greenland, parts of the shields in Brazil, and Africa) have almost no data. The raw average of all the continental heat flux values is $80\,\mathrm{mW\,m^{-2}}$ (Table 6, see also Pollack et al., 1993). However, there is a strong bias to high heat-flux values because many measurements were made in geothermal areas (e.g., the western United States and Baikal rift). In the United States, a large fraction of the more than 14 000 heat flux measurements belong to the Basin and Range Province. Excluding the values from the United States, the mean continental heat flux is only $66\,\mathrm{mW\,m^{-2}}$.

Bias in the sampling can be removed by area weighting the average as demonstrated in Table 6. Averaging over 1×1 windows yields a mean heat flux of $65.3\,\mathrm{mW\,m^{-2}}$. Using wider windows does not change this mean value significantly. The histograms of individual heat flux values and 1×1 averages have identical shapes, except for the extremely high values ($>200\,\mathrm{mW\,m^{-2}}$). Pollack et al. (1993) and later Davies and Davies (2010) had obtained a mean continental heat flux of $65–66\,\mathrm{mW\,m^{-2}}$ by binning heat flux values by tectonic age and weighting by the area. Different methods to estimate the mean continental heat flux consistently yield $63–66\,\mathrm{mW\,m^{-2}}$ (Jaupart and Mareschal, Chapter 6.05). For a mean continental heat flux value of $65\,\mathrm{mW\,m^{-2}}$, the contribution of all the continental areas (i.e., $210 \times 10^6\,\mathrm{km^2}$) to the energy loss of the Earth represents ≈ 14 TW. This number includes the submerged margins and continental areas with active tectonics,

Table 6 Continental heat flux statistics[a]

	$\mu(Q)$ (mW m^{-2})	$\sigma(Q)$ (mW m^{-2})	N(Q)
All measurements			
All values	70.1	152.0	35 065
Averages 1° × 1°	64.5	64.7	4261
Averages 2° × 2°	63.7	53.9	1955
Averages 3° × 3°	62.6	34.4	1205
Standard measurements only			
All values	80.3	215.1	17 256
Averages 1° × 1°	64.8	79.8	3391
Averages 2° × 2°	63.0	57.4	1673
Averages 3° × 3°	62.6	37.3	1051

[a] μ is the mean, σ is the standard deviation, and N is the number of values.

where higher than normal heat flux values are associated with thick radiogenic crust and shallow magmatic activity. Uncertainty in this number is due to lack of adequate data coverage in Greenland, Antarctica, and large parts of Africa. To estimate the induced uncertainty, we assume that heat flux in those areas is equal to either the lowest or the highest average heat flux recorded in well-sampled geologic provinces (36 and 100 mW m^{-2}, respectively). This procedure allows departures of ± 1.5 TW from the estimate of 14 TW. This uncertainty is certainly exaggerated because the poorly sampled regions are vast and contain geologic provinces of various ages and geologic histories; for instance, both Antarctica and Greenland are known to include high and low heat-flux regions. For the sake of simplicity, we retain a final uncertainty estimate of 1 TW.

7.06.4.2 Various Contributions to the Surface Heat Flux in Continental Areas

Determining the heat loss from the mantle through the continental lithosphere requires accounting for the radiogenic heat production. In stable continents, for ages greater than about 500 My, continents are near thermal steady state such that surface heat flux is the sum of heat production in the lithosphere and of the heat supply at the base of the lithosphere. The average heat flux does not vary significantly for provinces older than 500 My (Sclater et al., 1980) and, only in Archean (i.e., older than 2.5 Gy) provinces, it might be lower than in younger terranes (Morgan, 1983). We shall focus on estimating the crustal heat production and shall discuss briefly the contribution of the lithospheric mantle. The number of heat flux determinations in Archean and Precambrian provinces has increased during the past 30 years. With adequate sampling of heat flux and heat production, and detailed information on geology and crustal structure, the crustal and mantle components of the heat flux can now be determined. Further constraints are provided by (P, T) data on mantle xenoliths brought to the surface by kimberlite eruptions. It will be shown that, for stable regions, the crustal heat production makes the dominant contribution and the heat flux from the mantle is low.

Recently active regions are in a transient thermal regime and the high surface heat flux reflects cooling of the continental lithosphere. After removing the crustal heat production (which has been determined in stable provinces), it is possible to estimate the transient component of the heat flux, which originates in mantle cooling.

7.06.4.3 Estimating Moho Heat Flux

Many authors have assumed that the mantle heat flux is ≈ 25 mW m^{-2} in stable continental regions, because this was the lowest measured value (Cermak and Bodri, 1986; Pollack and Chapman, 1977a). As the number of measurements increased, it became clear that this cannot be true. Heat flux values ≈ 20 mW m^{-2} or less have been obtained in several different locations (Chapman and Pollack, 1974; Duchkov, 1991; Mareschal et al., 2000a,b, 2005; Swanberg et al., 1974), and the average heat flux over wide areas (500×500 km^2) of the Baltic and Siberian Shields is <18 mW m^{-2}. For the mantle heat flux to be equal to that at the surface, the whole crust below each measurement site must be completely devoid of heat-producing elements over a large area, which is not realistic. Tightening the range of values for the mantle heat flux was achieved in various ways.

7.06.4.3.1 Relationship between heat flow and heat production

Early attempts to calculate mantle heat flux relied on an empirical relationship between heat flux and heat production rate (Birch et al., 1968; Roy et al., 1968):

$$Q = Q_r + D \times H \qquad [44]$$

where Q is the local surface heat flux, H is the local surface heat production, and D is a length scale related to the thickness of a shallow layer enriched in radiogenic elements. The intercept Q_r represents the contribution of the mantle and crust below the enriched shallow layer. It was suggested that crustal heat production decreases exponentially as a function of depth down to the Moho (Lachenbruch, 1970). If this were true, values for D (≈ 10 km) would imply that the mantle heat flux is equal to Q_r. Although it was soon realized that it cannot be so, this is still taken for granted by many authors.

The significance of the empirical heat flow heat production relationship has been questioned on various grounds (England et al., 1980; Jaupart, 1983). For the rather small wavelengths involved, theory shows that the surface heat flux is only sensitive to shallow heat production contrasts (Jaupart, 1983; Vasseur and Singh, 1986). With more data available, it was found that the linear relationship is not verified in many geologic provinces (Jaupart and Mareschal, 1999; Jaupart et al., 1982). The crustal component of the heat flux can now be estimated from systematic investigations of lower crustal rocks, from both large granulite facies terrains (Ashwal et al., 1987; Fountain and Salisbury, 1981; Fountain et al., 1987) and xenolith suites (Rudnick and Fountain, 1995). The heat production values, obtained on samples from large exposure of granulite facies terranes in different areas of the Superior Province, are very consistent (≈ 0.4 μW m^{-3}). They appear to be representative of all granulite facies terranes worldwide, including the Ivrea zone (Joeleht and Kukkonen, 1998; Pinet and Jaupart, 1987). Thirdly, sampling in superdeep holes (Kola, Russia, and KTB, Germany) demonstrates that heat production shows no systematic variation with depth as would be required by the linear relationship. At Kola, the Proterozoic supracrustal rocks (above 4 km depth) have lower heat production (0.4 μW m^{-3}) than the Archean basement (1.47 μW m^{-3}) (Kremenentsky et al., 1989). At KTB, heat production decreases with depth at shallow levels, reaches a minimum between 3 and 8 km, and increases again in the deepest parts of the borehole (Clauser et al., 1997). Over a larger depth extent, studies of exposed crustal sections suggest a general trend of decreasing heat production with depth, but this trend is not a monotonic function (Ashwal et al., 1987; Fountain et al., 1987; Ketcham, 1996). Even for the Sierra Nevada Batholith where the exponential model had initially been proposed, recent research has shown that the heat production does not decrease exponentially with depth (Brady et al., 2006). In the Sierra Nevada, heat production first increases, then decreases, and remains constant in the lower crust beneath 15 km.

Another approach was to assume that the mantle heat flux is roughly proportional to the average surface heat flux (in a

proportion of ≈40%, Pollack and Chapman (1977b)). This analysis, however, was based on a small data set and implies that the crustal heat production is proportional to the mantle heat flux, which has no physical basis. In order to detect changes of mantle heat flux, small-scale heat flux variations are of little use because they record shallow heat production contrasts and one must work at a minimum scale of about 250 km (Mareschal and Jaupart, 2004). This places stringent constraints on data coverage because heat flux and heat production vary on a typical scale of 10 km due to the heterogeneity of the crust. In North America, there are sufficient data to illustrate the influence of the horizontal scale on heat flux and heat production. With individual measurements in a single geologic province, there is no meaningful (Q, H) relationship. At the scale of $\approx 250 \times 250$ km windows, a relationship between heat flux and heat production begins to emerge (Lévy et al., 2010). At the larger scale of geologic provinces with different crustal structures and histories ($\approx 500 \times 500$ km), the relationship is well defined and statistically significant (Figure 8). The data are close to a relationship of the form

$$\overline{Q} = Q_i + D\overline{H} \qquad [45]$$

where $\overline{Q}$ and $\overline{H}$ are province-wide averaged heat flux and heat production. That this relationship takes the same form as the 'local' relationship (eqn [44]) is probably fortuitous. In northern America, the latter is only valid for relatively small-scale variations (typically 10–50 km) of heat flux and heat production over Appalachian plutons and does not hold in the Precambrian provinces (Grenville, Trans-Hudson orogen, and Superior Province). The new relationship (eqn [45]) reflects variations of average heat flux on a much larger scale (>500 km) and relies on a very large data set. It implies that the average heat flux takes the same value Q_i at some intermediate crustal depth in all provinces. Formally, it is not possible to rule out variations of mantle heat flux between the five provinces, but the data require that such variations are exactly compensated by opposite variations of lower crustal heat production. It is hard to explain how this may be achieved in practice, and the most sensible hypothesis is that the mantle heat flux is approximately the same beneath the five provinces. For these provinces, independent geophysical and petrologic constraints on crustal structure show indeed that changes of crustal heat production account for the observed heat flux variations (Mareschal et al., 1999; Pinet et al., 1991).

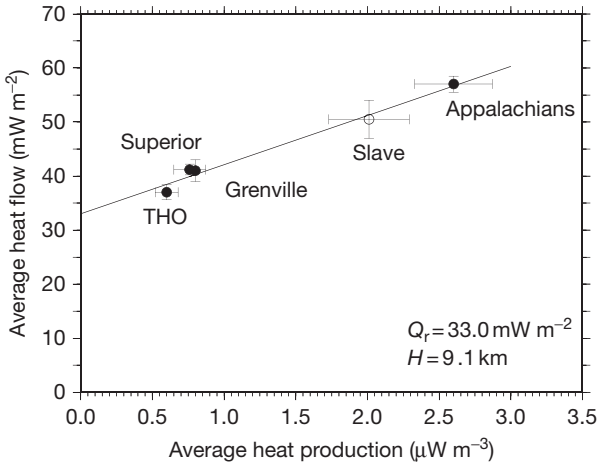


Figure 8 Averaged heat flux versus average surface heat production for five major geologic provinces of North America. The solid line is the best-fit linear relationship of eqn [45].

We now describe several methods for calculating crustal heat production, which consistently imply that the mantle heat flux is ≤ 18 mW m^{-2}.

7.06.4.3.2 Crustal heat production and Moho heat flux

In several parts of the Canadian Shield, heat flux values as low as 22 mW m^{-2} have been measured (Jaupart and Mareschal, 1999; Mareschal et al., 2000a,b). Similar or even lower values have also been reported for the Siberian Shield (Duchkov, 1991), the Norwegian Shield (Swanberg et al., 1974), and Western Australia (Cull, 1991). These low values, which correspond to areas where the crustal contribution is the lowest, provide an upper bound to the mantle heat flux. One may refine this estimate further by subtracting some lower bound for crustal heat production. Surface heat flux records a large-scale average of heat production, and one should consider a representative crustal assemblage and not a single rock type such as gabbro. For no crustal material are heat production estimates lower than 0.1 μW m^{-3} (Joeleht and Kukkonen, 1998; Pinet and Jaupart, 1987; Rudnick and Fountain, 1995). Over an average thickness of ≈40 km, the contribution of the crust must be at least 4 mW m^{-2}, and the mantle heat flux must be <18 mW m^{-1}. In Norway, Swanberg et al. (1974) obtained a heat flux value of 21 mW m^{-2} over an anorthosite body; after estimating the crustal heat production, they concluded that mantle heat flux is about 11 mW m^{-2}. The same value was obtained from the analysis of all the heat flux and radiogenic heat production data in the Norwegian Shield (Pinet and Jaupart, 1987).

A lower bound on mantle heat flux can be obtained by requiring that melting conditions are not attained in the crust in the absence of tectonic events and magmatic intrusions (Rolandone et al., 2002). In high heat-flux areas of the Canadian Shield, crustal rocks are at high temperatures today and were still hotter in the past when radiogenic heat production was higher. The condition of thermal stability provides a lower bound of 11 mW m^{-2} on the mantle heat flux. Combining this result with the independent constraints derived from present-day heat flux values leads to a range of 11–18 mW m^{-2} for the mantle heat flux beneath the Canadian Shield. Arguments different from these have led to the same range of values in other Precambrian areas (Guillou-Frottier et al., 1995; Gupta et al., 1991; Jones, 1988).

In several regions of the world, a large fraction of the crustal column has been exposed by tectonic processes. Sampling of such exposed cross sections allows the determination of the vertical distribution of radiogenic elements. If heat flux and seismic data are also available, it is possible to determine the total crustal heat production. For the Kapuskasing structure in the Canadian Shield where the crustal contribution could be determined, the Moho heat flux was calculated to be 13 mW m^{-2} (Ashwal et al., 1987; Pinet et al., 1991). The average crustal heat production can also be estimated in provinces where all crustal levels can be found at the surface. In these provinces, systematic sampling will yield an estimate of the average bulk crustal heat production. In the Grenville

province of the Canadian Shield, the average crustal heat production was determined to be 0.65 μW m^{-3} for an average surface heat flux of 41 mW m^{-2}. This yields a Moho heat flux of 15 mW m^{-2} (Pinet et al., 1991). Similar results have been reported for other shields in the world, including South Africa (Nicolaysen et al., 1981) and India (Roy and Rao, 2000), and are listed in Table 7.

Other methods have combined heat flux with other geophysical data, mainly long-wavelength Bouguer gravity, to estimate changes in crustal composition. A search for all models consistent with all the available data, including gravity data and bounds on heat production rates for the various rock types involved, leads to a range of 7–15 mW m^{-2} for the mantle heat flux in eastern Canada (Guillou et al., 1994).

The estimates mentioned earlier were derived using local geophysical and heat production data in several provinces and rely on knowledge of crustal structure. Independent determinations of the mantle heat flux may be obtained by considering the lithosphere thickness determined by seismic and xenolith studies. Pressure and temperature estimates from mantle xenoliths may be combined to determine a best-fit geotherm consistent with heat transport by conduction. Mantle heat flux estimates obtained in this manner depend on the value assumed for thermal conductivity. Available estimates are consistent with those deduced from crustal models and are listed in Table 7.

The estimates of Table 7 come from Archean and Proterozoic cratons where heat flux values are generally low. Heat flux values tend to be larger in younger stable continental regions. For example, heat flux is higher (57 mW m^{-2}) in the Appalachians than in the Canadian Shield. The crust of the Appalachians contains many young granite intrusions with very high heat production (>3 μW m^{-3}). The elevated heat flux can be accounted for by the contribution of these granites and does not require mantle heat flux to be higher than in the shield (Mareschal et al., 2000a,b; Pinet et al., 1991). Throughout stable North America, including the Appalachians, variations of the mantle heat flux may not be exactly zero but must be less than departures from the best-fitting relationship (Figure 8), or about ±2 mW m^{-2}. This estimate is close to the intrinsic uncertainty of heat flux measurements (Jaupart and Mareschal, 1999).

Allowing for the uncertainties and requiring consistency with low heat-flux measurements, we retain the range of 15 ± 3 mW m^{-2} for the mantle heat flux in stable continents. For this range, the differences of average heat flux between geologic provinces cannot be accounted for by changes of mantle heat flux and must be attributed to changes of crustal heat production. The ranges of heat flux and heat production values are the same for all provinces between 200 My and 2.5 Gy, with a weak trend of decreasing average heat flux and heat production with age (Perry et al., 2006). The range is narrower in Archean provinces where high heat-flux values are not found, possibly because a very radioactive crust would have been too hot to be stabilized (Morgan, 1983). Averaging the heat production of the crust of different ages yields a range of 0.79–0.99 μW m^{-3} (Table 8 and Jaupart and Mareschal (2013)).

Table 7 Various estimates of the heat flux at Moho in stable continental regions

Location	*Heat flux (mW m^{-2})*	*Reference*
Norwegian Shield	11[a]	Swanberg et al. (1974) and Pinet and Jaupart (1987)
Baltic Shield	7–15[a]	Kukkonen and Peltonen (1999)
Siberian craton	10–12[a]	Duchkov (1991)
Dharwar Craton (India)	11[a]	Roy and Rao (2000)
Kapuskasing (Canadian Shield)	11–13[a]	Ashwal et al. (1987) and Pinet et al. (1991)
Grenville (Canadian Shield)	13[a]	Pinet et al. (1991)
Abitibi (Canadian Shield)	10–14[a]	Guillou et al. (1994)
Trans-Hudson orogen (Canadian Shield)	11–16[b]	Rolandone et al. (2002)
Slave province (Canada)	12–24[c]	Russell et al. (2001)
Vredefort (South Africa)	18[a]	Nicolaysen et al. (1981)
Kalahari Craton (South Africa)	17–25[c]	Rudnick and Nyblade (1999)

[a]Estimated from surface heat flux and crustal heat production.
[b]Estimated from condition of no melting in the lower crust at the time of stabilization.
[c]Estimated from geothermobarometry on mantle xenoliths.

7.06.4.4 Recently Active Regions and Continental Margins

Submerged and recently active (i.e., during the past 200 My) continental areas cover 92 × 10^6 km^2, ≈45% of the total continental surface (Table 5). These regions are not in thermal steady state and are characterized by higher heat flux than the continental average. Because of the long thermal relaxation time of the continental lithosphere, present surface heat flux includes the inputs of heat from the mantle of the past 100–200 My. The crustal component can now be calculated from crustal thickness and average heat production. After accounting for crustal heat production, the heat from the mantle (including the transient component) can be estimated.

7.06.4.4.1 Compressional orogens

In compressional orogens, crustal and lithospheric thickening result in reduced temperature gradients and heat flux, but the total heat production in the thick crust is high. These two competing effects lead to a complex transient thermal structure, and few generalizations can be made on the surface

Table 8 Estimates of bulk continental crust heat production from heat flux data (Jaupart and Mareschal, 2013)

Age group	*Heat production (μW m^{-3})*	*Total (40 km crust) (mW m^{-2})*	*% Area* [a]
Archean	0.56–0.73	23–30	9
Proterozoic	0.73–0.90	30–37	56
Phanerozoic	0.95–1.21	37–47	35
Total continents	0.79–0.99	32–40	

[a]Fraction of total continental surface, from model 2 in Rudnick and Fountain (1995).

heat flux. For instance, very high heat-flux values (>100 mW m^{-2}) have been measured on the Tibetan Plateau (Francheteau et al., 1984; Hu et al., 2000; Jaupart et al., 1985). They have been attributed to shallow magma intrusions and yield little information on the mantle heat flux. In contrast, present surface heat flux remains low in the Alps, and after removing the crustal heat production, heat flux at the Moho is estimated to be as low as 5 mW m^{-2} (Vosteen et al., 2003). After removing the crustal contribution, heat flux from the mantle is also low beneath the North American Cordillera (Brady et al., 2006) and beneath the South American Cordillera, at least where it has not been affected by back-arc extension (Henry and Pollack, 1988).

7.06.4.4.2 Rifts and continental zones of extension

In rifts, recently extended regions, continental margins, and basins, heat flux is higher than in stable areas because of a large transient component, which ultimately represents additional outputs of heat from the mantle. Crustal extension and lithospheric thinning will instantly result in steepening the temperature gradient and increasing the heat flux. Thermal relaxation from the initial conditions depends on the boundary condition at the base of the lithosphere.

In zones of extension and continental rifts, heat flux values are significantly higher (75–125 mW m^{-2}) than in stable regions (Morgan, 1983). A striking feature of extension zones is that the transition between the region of elevated heat flux and the surrounding is as sharp as the sampling allows one to determine, that is, Colorado Plateau–Basin and Range Province in North America (Bodell and Chapman, 1982), East African Rift–Tanzania Craton (Nyblade, 1997), and Baikal rift–Siberian craton (Poort and Klerkx, 2004). This suggests that the enhanced heat flux is not due to conductive processes but is the direct result of extension and lithospheric thinning. With sufficient sampling, one can detect short-wavelength variations of heat flux that can be attributed to the cooling of shallow magmatic intrusions and groundwater movement. The actual heat loss is higher than the average conductive value because of heat transport by hot springs and magmas. Lachenbruch and Sass (1978) had estimated that the heat delivered by volcanic systems in rifts and in the Basin and Range Province is negligible. They had also argued that the integrated effect of heat transport by groundwater is small for this province, with the exception of the Yellowstone system, where locally the heat flux is >40 W m^{-2}. Heat loss for the entire Yellowstone system has been estimated to be ≈ 5 GW, including both conductive and convective components (Fournier, 1989). It would thus require 200 'Yellowstones' to increase the continental heat loss by 1 TW. The effect of continental hot spots on the budget seems presently negligible. Estimates for the total heat loss through geothermal systems in the East African Rift are comparable to those of Yellowstone (Crane and O'Connell, 1983). Similar values have been inferred for Baikal (Poort and Klerkx, 2004). In continental as well as in oceanic rifts, the heat loss is underestimated because of hydrothermal heat transport. However, because continental rifts are narrow and their total surface area is small, they account for a small fraction of the total continental heat loss, and errors on their contribution will have a negligible impact on the end result.

The contribution of wide regions of extension is more significant than that of rifts. In the Basin and Range Province in the southwestern United States, the high average heat flux (105 mW m^{-2}) has been attributed to an extension of 100% (Lachenbruch and Sass, 1978). This interpretation depends on assumptions on the preextensional heat flux and on crustal heat production. Early estimates of the mantle heat loss are probably too large because the crustal heat production was underestimated (Ketcham, 1996). It now appears that the average heat production of the crust is the same as in stable regions and yields a total crustal heat flux contribution of ≈ 33 mW m^{-2}. This implies that the transient component due to cooling and the mantle heat flux in the Basin and Range Province add up to a total of ≈ 70 mW m^{-2} (Ketcham, 1996). Physical models to account for this heat loss assume either delamination of the lithospheric mantle or stretching and transport of heat into the lithosphere by magmatic intrusions (Lachenbruch and Sass, 1978). Regardless of the mechanism, at least 2/3 of the heat flux in regions of extension comes from the mantle.

Large igneous provinces testify of periods of enhanced volcanic activity in the continents. Their effect on the heat flow is however negligible. In the Deccan, where 500 000 km^3 of basalts was deposited *c.*60 My, there is no heat flow anomaly, suggesting that the magmas did not heat up the lithosphere over large volumes. Assuming that the lavas were deposited in 1 My, the heat that they carried to the surface contributed <0.1 TW to the energy budget.

7.06.4.4.3 Continental margins: Passive

Continental margins are the transition zones between ocean and continents and have the characteristics of both: crust that is enriched in radioelements and rests on top of thin lithosphere. For a total length of about 100 000 km (Bradley, 2008) and an average width of 300 km, they cover 30×10^6 km^2, or 15% of the total continental area. Passive margins form during the breakup of continents as continental lithosphere extends and seafloor spreading begins. They come in many different styles and shapes, depending on width, abundance of magmatic activity, and sediment thickness. Their deepest parts are generally made of hyperextended continental crust and in some cases exhumed continental mantle (Boillot et al., 1980; Unternehr et al., 2010; Whitmarsh et al., 2001). They subside as the thermal perturbation caused by thinning and rifting decays (Sleep, 1971; Vogt and Ostenso, 1967). Most of the present-day ones, which border the Atlantic and Indian Oceans, formed during the breakup of Gondwanaland and have returned to thermal equilibrium. A few recent ones (Red Sea, Gulf of Aden, North Atlantic, etc.) are in transient thermal regimes.

Heat flow measurements on passive margins are scarce because conventional techniques are not adapted to shallow water environments and/or to the high-porosity sedimentary layers. Many heat flow values have been deduced from temperatures recorded in oil wells on continental platforms or from the depth of the bottom simulating reflector (BSR) on continental slopes. Oil well data are either bottom hole temperatures, which are perturbed by mud circulation and drilling operations, or fluid temperatures during completion tests, which are considered to be more reliable. BSRs are strong

reflectors that occur at the lower stability limit of gas hydrates and free gas (Singh et al., 1993). Because the stability limit depends on pressure and temperature, the geothermal gradient can be estimated from the depth of the BSR. Usually, thermal conductivity is estimated empirically from lithology and porosity, which adds to the uncertainty. The accuracy of these data has been assessed in a few deep margins through a comparison with conventional marine measurements (Lucazeau et al., 2004). Analysis of the global data set shows that heat flow measurements are highly scattered and decrease with the age of the margin (**Figure 9**), as expected (e.g., McKenzie, 1978). The values, however, are higher than the predictions of conductive models of passive rifting with acceptable amounts of stretching (**Figure 9**). Local surveys show indeed that the surface heat flux is perturbed with respect to the adjacent continental and oceanic domains.

In young continental margins (Red Sea and Gulf of Aden), heat flow is higher than could be expected from the magnitude of continental crust extension (Lucazeau et al., 2008; Martínez and Cochran, 1989). Heat flow increases abruptly at the landward edge of the transition region between oceanic and continental domains (OCT). The sharp lateral variation and the magnitude of the heat flux variation both require the presence of a shallow thermal anomaly. In the Gulf of Aden, for example, heat flow increases from 60 to 120 mW m^{-2} over a few km and then remains almost constant for several tens of kilometers over the OCT (Lucazeau et al., 2008). Higher heat-flux values (up to 300 mW m^{-2}) have been measured in the northern Red Sea, where continental breakup is occurring today (Martínez and Cochran, 1989). Independent evidence for a shallow thermal anomaly comes from the depth of the seafloor. Sediments that are deposited in the OCT during early phases of extension and rifting belong to shallow water environments, indicating that the isostatic consequences of crustal thinning are compensated by the buoyancy of hot mantle (Huismans and Beaumont, 2011).

In mature continental margins, the surface heat flux is usually higher than expected from passive rifting models. In the Labrador margin (Canada), heat flux is about 15–25 mW m^{-2} higher than on the adjacent continent (Goutorbe et al., 2007; Mareschal et al., 2000a,b). This is observed in many other regions, for example, the Gulf of Guinea (Lucazeau et al., 2004), Senegal (Latil-Brun and Lucazeau, 1988), South Africa (Goutorbe et al., 2008a), the Gulf of Mexico (Husson et al., 2008; Nagihara and Opre, 2005), South West India and Makran (Calves et al., 2010; Rao et al., 2001), Mozambique margin (Nyblade, 1997), Australian margins (Goutorbe et al., 2008b), and the Iberia margin (Louden et al., 1997). Many different explanations have been put forward, including high heat production in the sediments (McKenna and Sharp, 1998) or in the stretched continental crust (Louden et al., 1997; Nagihara and Opre, 2005), volcanism (Rao et al., 2001),

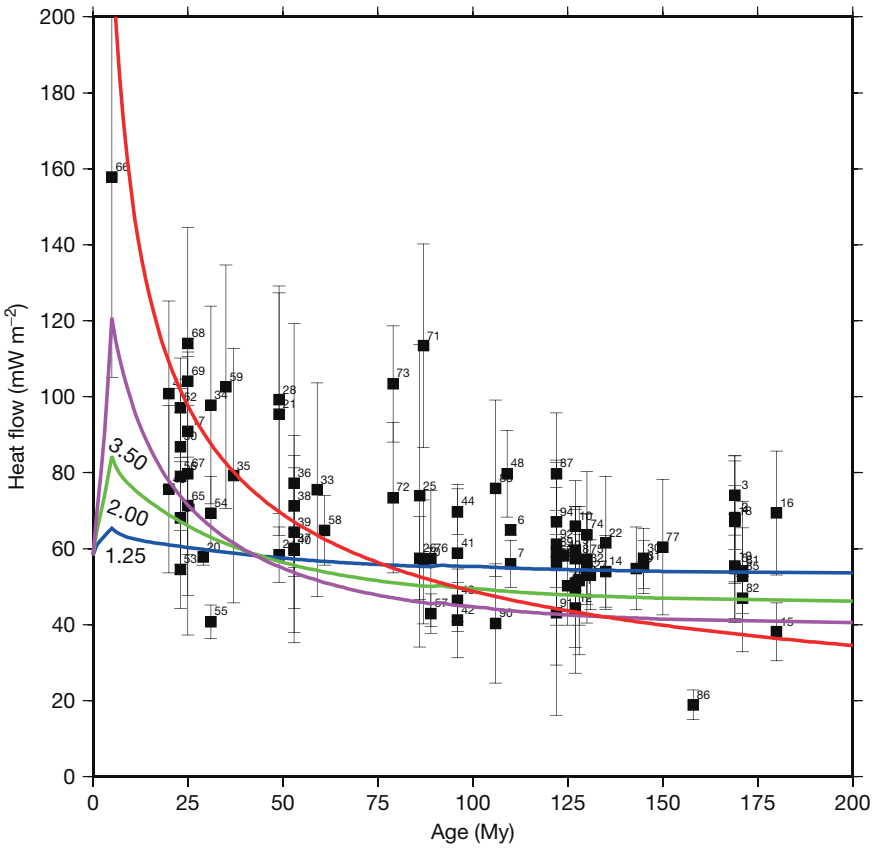


Figure 9 Heat flow versus continental margin age. Results of the homogeneous extension model (McKenzie, 1978) modified to account for crustal heat production for stretching factors = 1.25 (blue), 2.00 (green), and 3.50 (magenta). The dark red curve shows the predictions of the half-space cooling model with $C_Q = 490$.

erosion, renewed rifting (Husson et al., 2008), and fluid circulation (Deming et al., 1992; Reiter and Jessop, 1985). The systematic pattern of heat flux variations in continental margins, however, is best accounted for by a local mantle convective circulation (Armitage et al., 2013; Grigne et al., 2007; Guillou and Jaupart, 1995; King and Anderson, 1998).

Measurements on passive margins, as defined in the UNESCO Geological Map of the World (CCGM/CGMW, 2000), account for $<5\%$ of the entries in the global heat flow database. They average 62 mW m^{-2} (**Table 5**), which is close to the bulk continental average. OCTs are usually mapped as oceanic domains, but do not affect the statistics significantly if they are lumped with continental margins proper. Young rifted margins are associated with significantly higher heat flow values, but they account for $<10\%$ of the total passive margin area (Bradley, 2008). Multiplying the average heat flux by the area of passive margins, we obtain a heat loss contribution of 2.0 TW.

7.06.4.4.4 Continental margins: Active

Active continental margins lie at the transition between oceans and continents in subduction zones. There, the surface heat flux depends on the thermal structure and dip of the subducting plate (Molnar and England, 1990) and records local perturbations due to sedimentation (Hyndman et al., 1993; Marcaillou et al., 2006) and fluid flow (Henry and Wang, 1991; Hyndman et al., 1993). Theory predicts that the surface heat flux should decrease from the trench to the accretionary prism and forearc basin.

In forearc domains, heat flow reflects the thermal structure of the subducting plate and is related to the age of the plate. The average value is 70 mW m^{-2} (**Figure 10**). In back-arc basins, heat flow is controlled mostly by the magnitude of extension and magmatic activity and takes large values with an average of 91 mW m^{-2} (**Table 5**). Similar values are found in active mobile belts, such as the Canadian Cordillera and Cascadia (Hyndman, 2010; Lewis et al., 2003).

For a total area of 30×10^6 km^2, active margins contribute $\sim$2.4 TW to the total heat loss, assuming equal partitioning between forearcs and back-arcs.

7.06.4.5 Mantle Heat Loss Through Continental Areas

There is a major difference in the thermal regime between stable and active continental regions. In stable continental regions, the mean heat flux is low ($\leq$55 mW m^{-2}) and mostly comes from crustal heat production. Heat flux from the mantle is $\approx$15 mW m^{-2}. In extensional regions, the high heat flux ($\geq$75 mW m^{-2}) includes contributions from crustal radioactivity ($\approx$30 mW m^{-2}) as well as from thermal relaxation of the mantle. Despite their thin crust, continental margins also have higher than average heat flux ($\approx$80 mW m^{-2}) because they are cooling after being extended.

Different methods lead to a value of 14 TW for the integrated heat flux from continental areas. Neglecting geothermal and volcanic transport has no significant impact on this value. The estimated average heat production of the continental crust ranges between 0.79 and 0.99 μW m^{-3} (Jaupart and Mareschal, 2013), and the total volume of continental crust is $\approx0.73\times10^{10}$ km^3, which gives a total heat production in the crust between 6 and 7 TW.

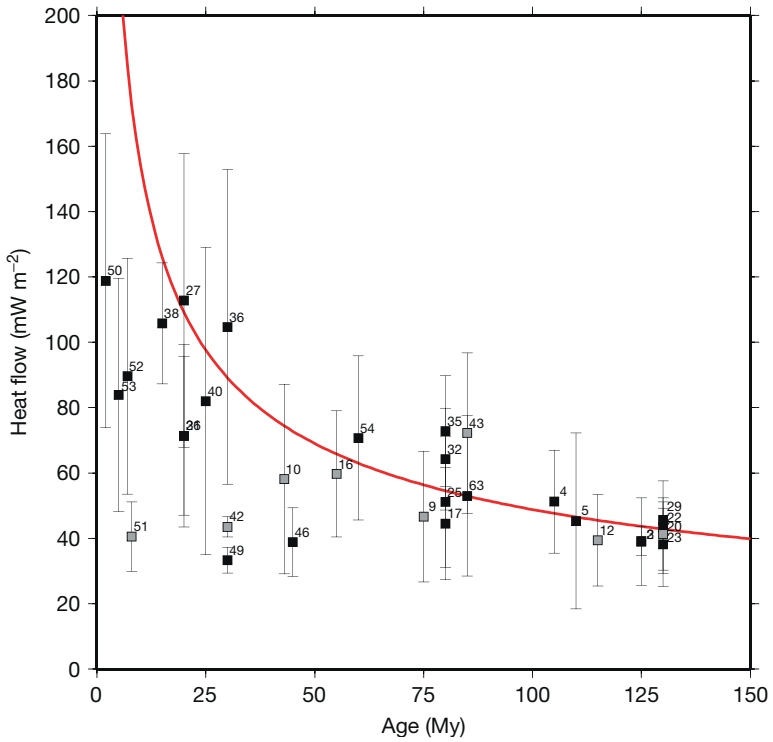


Figure 10 Heat flow measured on trenches and forearc basins versus age of the subducting plate. Black symbols are measurements close to the trenches and gray symbols are measurements obtained at distance >100 km. Oceanic half-space cooling model for $C_Q=490$ is also shown (red curve).

Little is known about the amounts of radiogenic elements in the lithospheric mantle. Direct estimates rely on a few exposures of peridotite massifs, which are typically depleted (Rudnick et al., 1998), and on mantle xenoliths from kimberlite pipes, which are usually enriched (Russell et al., 2001). Considerations on the thermal stability of continental roots and consistency with heat flux measurements as well as with petrologic temperature estimates lead to the conclusion that enrichment must be recent and associated with metasomatic infiltrations (Jaupart and Mareschal, 1999; Russell et al., 2001). This enrichment process is probably limited in both area and volume and our best estimate of radiogenic heat production in the lithospheric mantle comes from peridotite massifs. For the sake of completeness, we take a value of 0.02 $\mu W\ m^{-3}$ (Rudnick et al., 1998) and consider an average lithosphere thickness of 150 km. The total heat thus generated in the subcontinental lithospheric mantle is about 0.5 TW, which is only accurate within a factor of about 2.

Subtracting the contribution of radioactive sources from the total heat loss out of continents, we thus arrive at an estimate of the heat input from the mantle of 6–7 TW, about half of which is brought through the tectonically active regions and the continental margins.

7.06.4.6 Summary

Heat flux data are now available for provinces of all ages, including Archean cratons that were poorly sampled 30 years ago. About half of the heat loss through continents is accounted for by crustal radiogenic heat production. Stable continents allow a small heat flux of about 15 $mW\ m^{-2}$ out of the convecting mantle and act as insulators at the surface of the Earth.

7.06.5 Heat Sources

7.06.5.1 Radiogenic Sources in the Bulk Silicate Earth

The composition of our planet cannot be measured directly for lack of direct samples from the lower mantle and the core and has been estimated using various methods. It had been noted by Birch (1965) that if the Earth had a chondritic composition, its heat production would match what was then thought to be the heat loss (30 TW). This remarkable coincidence did not resist close scrutiny. It was soon noted that the Earth is depleted in K relative to chondrites and this reduces the heat production (Wasserburg et al., 1964). With the same amount of U and Th as in the chondrites, and a terrestrial K/U ratio, the total heat production was estimated to be only 20 TW (Wasserburg et al., 1964). On the other side of the balance, the heat loss is now believed to be much larger than it was then.

All attempts to construct a bulk Earth composition model rely on two different kinds of samples: meteorites, which represent the starting material, and pieces of today's upper mantle. Both show rather extensive variations of composition due to their different histories. Processes in the early solar nebula at high temperature contribute one type of compositional variation. Processes within the Earth, which occur at lower temperatures, contribute another type of compositional variation. Stated schematically, one has a range of compositions from the early solar system and a range of compositions for the upper mantle of the Earth, and one must devise a procedure to account for the different processes that have affected the two types of samples.

Chondrites represent samples of undifferentiated silicate material from the solar system prior to melting and metallic core segregation. Their composition derives from the solar composition altered by processes in the early solar nebula that have generated different families of chondrites. Perturbations are essentially brought in the gas state and elemental behavior is classified according to volatility (or condensation temperature). For our present purposes, the important elements are uranium, thorium, and potassium. The first two are associated with very high condensation temperatures and called 'refractory lithophile' elements (RLE). That these two elements have the same behavior in the early solar system is demonstrated by the fact that they have the same ratio in all types of chondritic meteorites. Potassium is a 'moderately volatile' element with a lower condensation temperature. The best match with solar concentration ratios is achieved by CI chondrites, which explains why many Earth models have relied on them. However, we do know that CI chondrites have larger amounts of volatiles, including water and CO_2, than the Earth. As regards samples from the Earth's mantle, one may establish a systematic compositional trend through the samples and identify the most primitive (and least differentiated) material. With these problems in mind, we review the four main types of approaches that have been used and the resulting estimates of Earth composition. We shall refer to the bulk silicate Earth (BSE) that corresponds to mantle and crust.

The first method relies on direct samples from the mantle. Ringwood (1962) argued that basalts and peridotites are complementary rocks, such that the latter is the solid residue of the partial melting event that led to basalt genesis and extraction. Thus, mixing them back together with the appropriate proportions yields the starting material, which was named 'pyrolite.' Clearly, one has to choose the samples that have not been affected by leaching and low-temperature alteration. Unfortunately, this procedure is not efficient for uranium, which is very mobile and seldom resists weathering.

A second method relies on a choice for the starting material. BSE concentrations are obtained by working one's way through the processes that turn meteorites into Earth-like material: devolatilization (loss of water, CO_2, and other volatile elements present in very small amounts in the Earth), followed by reduction (loss of oxygen) and segregation of a metallic core. Errors associated with this obviously come not only from the mass loss estimates but also from the starting chondrite composition since the different groups of meteorites are all quite heterogeneous. Many authors (e.g., Hart and Zindler, 1986) have used CI chondrites, as explained earlier. Javoy (1995, 1999) had argued in favor of a different type of meteorites. His line of reasoning focuses on the oxidation state of the solar nebula when it started to condense. The only meteorites with the right isotopic ratios and oxidation state are enstatite chondrites (EH) (Javoy et al., 2010). These chondrites are largely degassed, save for sulfur, so that the volatile loss correction is small. RLE elements are not affected by devolatilization and core segregation and one can use this property to avoid forward calculations that are error-prone given the many variables and parameters involved. Calcium belongs to the RLE group and

the Th/Ca and U/Ca ratios of EH chondrites, which are necessarily identical to the BSE values, are known (Wasson and Kalleymen, 1988). In the model of Kaminski and Javoy (2013) for the formation of the Earth's mantle and core out of EH chondritic material, the BSE Ca concentration is 1.71 ± 0.08 wt%, which leads to values of 18.1 and 60.2 ppb for U and Th, respectively.

The debate about the starting material for the Earth is in full swing at the time of writing, and one can expect further developments as new isotopic measurements become available. The choice of starting material has implications for studies of the early solar nebula and meteorites on the one hand and of the Earth's mantle structure on the other hand. For example, one argument against EH chondrites as Earth building material is that their Si/Mg ratio is much larger than that of the upper mantle, which requires a lower mantle enriched in Si (Fitoussi and Bourdon, 2012). In fact, it may be that the Earth was not made out of chondritic material. This is suggested by differences in the ^{142}Nd concentrations and ^{142}Nd/^{144}Nd ratios of chondrites and terrestrial rocks (Boyet and Carlson, 2005; Boyet and Gannoun, 2013; Caro et al., 2008). Using the terrestrial Nd values and geochemical systematics, Caro and Bourdon (2010) and Jackson and Jellinek (2013) had built a nonchondritic BSE model that has lower U and Th concentrations than chondritic ones (**Table 9**).

A third method tries to avoid a specific choice for the starting composition and aims at determining it. Hart and Zindler (1986) determined the compositional trends of chondritic meteorites and peridotites, which are not parallel to one another. Each trend records the effects of the two different sets of processes operating in the primitive solar nebula and in the Earth, and their intersection can only be the starting Earth material. In this case, the error comes from the scatter around the two compositional trends.

A fourth method relies on elemental ratios. For RLE, such as uranium and thorium, the concentration ratio is independent of chondrite type and is a property of the starting Earth material. Once these ratios have been determined, two procedures can be used to determine primitive abundances from measurements on peridotite samples. In one procedure, one starts with one specific element for which one can determine a reliable value for the bulk Earth and work sequentially to all the others using elemental ratios. The element of choice is Mg because, although it is not the most refractory element, its behavior during melting and alteration is well understood (Palme and O'Neill, 2003). The other procedure is to study the relationship between abundance and elemental ratios (McDonough and Sun, 1995). Changes in both values can be accounted for by variations in the degree of melting of peridotite. McDonough and Sun (1995) used a linear regression to extrapolate the data array to the chondritic ratio, which gives the primitive abundance. Depletion processes due to melt extraction are intrinsically nonlinear, however. With a more realistic treatment of these effects and tighter statistical analysis of the highly scattered data, Lyubetskaya and Korenaga (2007) had obtained a model for the BSE that is more depleted than previous ones. Uncertainties on the uranium, thorium, and potassium concentrations are large ($\approx$15%; see **Table 9**).

As regards the radioactive elements of interest here, the 'pyrolite' method is unreliable and was not used. **Table 9** lists

Table 9 Radio-element concentration and heat production in meteorites, in the Bulk Silicate Earth, in Earth mantle and crust

	U (ppm)	*Th (ppm)*	*K (ppm)*	A* *(pW kg^{-1})*
CI chondrites				
Palme and O'Neill (2003)	0.008	0.030	544	3.5
McDonough and Sun (1995)	0.007	0.029	550	3.4
EH chondrites				
Wasson and Kalleymen (1988)	0.009	0.030	/	
Bulk silicate Earth				
From CI chondrites				
Javoy (1999)	0.020	0.069	270	4.6
From EH chondrites				
Kaminski and Javoy (2013) [b]	0.018 ± 0.001	0.060 ± 0.003	217 ± 11[a]	4.0 ± 0.2
From chondrites and lherzolites trends				
Hart and Zindler (1986)	0.021	0.079	264	4.9
From elemental ratios and refractory lithophile element abundances				
McDonough and Sun (1995)	$0.020 \pm 20\%$	$0.079 \pm 15\%$	$240 \pm 20\%$	4.8 ± 0.8
Palme and O'Neill (2003)	$0.022 \pm 15\%$	$0.083 \pm 15\%$	$261 \pm 15\%$	5.1 ± 0.8
Lyubetskaya and Korenaga (2007)	0.017 ± 0.003	0.063 ± 0.011	190 ± 40	3.9 ± 0.7
Jackson and Jellinek (2013)	0.014 ± 0.003	0.055 ± 0.011	166 ± 30	3.4 ± 0.5
Depleted MORB source				
Workman and Hart (2005)	0.0032	0.0079	25	0.59
Average MORB mantle source				
Su (2000) and Langmuir et al. (2005)	0.013	0.040	160	2.8
Peridotites	0.006	0.02	100	1.5
Continental crust				
Rudnick and Gao (2003)	1.3	5.6	1.5×10^4	330
Jaupart and Mareschal (2013)	/	/	/	293–352

[a]Calculated from the U value and K/U = 12 000.

[b]U and Th values deduced from the Ca concentration and the chondritic U/Ca and Th/Ca ratios.

estimates obtained by different authors using the other methods. One should note that concentration ratios are constrained more tightly than absolute concentrations. One cannot separate uncertainties due to the starting chemical data from those of the calculation algorithm, because each author uses his own data and method. Without major improvement in our understanding of early planetary accretion, it may well prove impossible to reduce the spread of results.

Values for U, Th, and K concentrations yield estimates of the radiogenic heat production rate in the Earth. We have used the revised decay constants listed in **Table 10** (Dye, 2012). Those differ slightly (<4%) from the earlier estimates given by Rybach (1988), which are commonly used in the geophysical literature. Heat production values vary within a restricted range, from 3.4 to 5.1 pW kg^{-1}. The BSE model of Palme and O'Neill (2003) leads to the highest value and may well be an overestimate. The EH chondrite model does not lead to a major difference for the present-day heat production rate but implies important changes at early ages because its Th/U and K/U ratios differ strongly from the others. In spite of the differences, the range of compositions from the various estimates (≈15%) is sufficiently wide to allow some overlap between them.

The different models for the bulk silicate Earth, which includes continental crust, are listed in **Table 9**. They lead to a total rate of heat production of 18 TW, with an uncertainty of 28%. After removing the contributions of the continental crust (6–7 TW) and the lithospheric mantle (≈1 TW), heat production in the mantle amounts to a total of 11 TW, with an uncertainty of 55%.

Independent constraints can be derived from the compositions of MORBs, which have been comprehensively sampled. MORBs span a rather wide compositional range and common practice has been to define end-members associated with different chemical reservoirs. According to this framework, depleted MORBs come from a depleted reservoir whose complement is enriched continental crust. Enriched basalts are attributed to primitive mantle tapped by deep mantle plumes or to secondary processes, for example, infiltrations of low-degree melts and metasomatic fluids in subduction zones (Donnelly et al., 2004). The heat production rate of the depleted MORB mantle source is ≈0.6 pW kg^{-1} (i.e., at this rate, the entire mantle would generate only 2.4 TW). This source, however, does not provide an exact complement of average continental crust (Workman and Hart, 2005). An alternative approach avoids the separation of different mantle reservoirs and determines the average composition of all the mantle that gets tapped by mid-ocean ridges (Langmuir et al., 2005; Su, 2000). Composition of the average MORB mantle source is then derived from a well-constrained melting model. This mantle reservoir is a mixture of different components and is the average mantle lying below oceanic ridges. It is depleted in compatible elements and may be interpreted as the mantle reservoir that has been processed to form continents (Langmuir et al., 2005). There may be a volume of primitive mantle lying at depth that has never been sampled by mid-oceanic ridges, however. Assuming that the average MORB source extends through the whole mantle leads to a lower bound of 11 TW on the total mantle heat production (**Table 9**). Adding radioelements from the continental crust and lithospheric mantle, which contribute 7–8 TW, we obtain a lower bound of 18 TW for the total rate of heat production in the Earth. This is consistent with the BSE models and their uncertainties.

7.06.5.2 Geoneutrino Measurements

So far, it has been impossible to directly detect heat-producing elements in the mantle and measure their concentration, but this may change with the advent of 'geoneutrino' detection. 'Geoneutrinos' are antineutrinos generated by the decay of the radioelements in the Earth and they can be captured by the large liquid scintillation detectors of underground neutrino observatories (Dye, 2010; Enomoto et al., 2007; Fiorentini et al., 2005; Kamland collaboration, 2011; Raghavan et al., 1998; Rothschild et al., 1998). With the present technology, only geoneutrinos from U and Th decay can be detected, but geoneutrinos from K decay cannot because their energy is below the detection threshold. With the very small cross section of the geoneutrinos (on the order of $10^{-44}cm^2$), very long exposure times in large detectors are required for statistically significant measurements. The geoneutrino flux integrates all the sources in the Earth, from the crust as well as from the mantle. In a continental observatory, at least 75% of this flux comes from the crust (Enomoto et al., 2007). For determining the geoneutrino flux from the mantle, the local crustal contribution must be very accurately accounted for, which represents a difficult challenge (Dye, 2010; Perry et al., 2009). This problem could be circumvented by deploying large submarine detectors on the seafloor where the crustal contribution to the geoneutrino flux is minimum (Dye et al., 2006). An alternative solution would be to measure the direction of the interacting geoneutrinos, which

Table 10 Heat production constants

Isotope/element	*Natural abundance (%)*	*Half-life (year)*	*Energy per atom (pJ)*	*Heat production per unit mass of isotope/element ($\mu W\ kg^{-1}$)*
^{238}U	99.27	4.46×10^9	7.65	95.13
^{235}U	0.72	7.04×10^8	7.11	568.47
U				98.5
^{232}Th	100	1.40×10^{10}	6.48	26.3
Th				26.3
^{40}K	0.0117	1.26×10^9	0.110	28.47
K				3.33×10^{-3}

The table presents recent determinations that account for the energy taken away by neutrinos (Dye, 2012). These values differ slightly from the values commonly used by geoscientists (e.g., Rybach, 1988).

would also place constraints on the radial distribution of the mantle sources (Hochmuth et al., 2007; Mareschal et al., 2012).

In recent years, two groups have reported on their geoneutrino observations. The Kamioka Underground Observatory in Japan has reported on measurements during 9 years of operation. There, the geoneutrino flux is 39 ± 10 terrestrial neutrino units (1 TNU amounts to 1 geoneutrino event per 10^{32} protons per year exposure time) (Kamland collaboration, 2011). A crustal contribution to 27 ± 11 TNU was inferred from a model of the Japanese crust, leaving 12 TNU from the mantle. The 'noise-free' observations made after the shutdown of the Japanese nuclear reactors following the Fukushima accident have provided new estimates of the signal rate at Kamioka. The most recent analysis gives a signal rate of 30 ± 7 TNU, which leads to ≈ 7 TNU from the mantle (Kamland Collaboration (Gando, A., and 45 Collaborators), 2013).

The latest results from the much smaller detector at Borexino, Italy, yield a higher signal rate of 39 ± 12 TNU with a crustal component now estimated at 23 TNU (Borexino Collaboration Group (Bellini and 88 Collaborators), 2013). Combining these results leads to a total heat production for uranium and thorium of 14 ± 8 TW (**Table 11**). Using an average K/U ratio of 12 000, we add heat generated by potassium and obtain a total heat production of 18 ± 9 TW. The difference between the Kamioka and the Borexino values is smaller than previously reported but leaves a large error bar on the mantle contribution. The error is due both to the experimental uncertainty and to the uncertainty on the crustal contribution. These results are remarkable because the observatories were built for particle physics in sites poorly suited to accurately calculate the neutrino flux from a very heterogeneous crust. Neutrino geophysics is still in its infancy, but the results obtained so far give confidence that, with longer exposure times and better statistics in well-located observatories, we shall get robust constraints on the concentration of radioactive elements in the mantle.

7.06.5.3 Heat Flux from the Core

The outer core is made of molten iron with very low viscosity and loses heat to the much more viscous deep mantle. Thus, the heat flux out of the core is controlled by the efficacy of mantle convection and cannot be considered as an independent input. Nevertheless, the thermal evolution of the core determines the energy available to drive the geodynamo, which can be constrained from thermodynamics. This question is briefly covered here and the interested reader should consult the volume concerning the core (see in particular Nimmo, **Chapter 8.02**) and some standard references (Braginsky and Roberts, 1995; Gubbins and Roberts, 1987; Labrosse, 2005a,b; Lister and Buffett, 1995), as well as Hernlund and McNamara (**Chapter 7.11**).

An energy balance can be written for the core, in much the same way as for the mantle. The main differences come from electromagnetic processes and chemical buoyancy due to inner core crystallization. The low viscosity maintains the convective state very close to the reference (radially symmetrical) state. The convective velocity at the surface of the core, which is of the order of $10^{-4}\,\mathrm{m\,s^{-1}}$ (Hulot et al., 2002; see also Jackson and Finlay, **Chapter 5.05**), suggests relative density fluctuations of order 10^{-9} (Braginsky and Roberts, 1995; Labrosse et al., 1997). Such fluctuations correspond to temperature variations $\delta T \sim 10^{-4}\,\mathrm{K}$ in the absence of other effects. This is much smaller than the secular temperature decrease, implying separation of scales between the upper boundary layer and the interior of the outer core. Thus, for the secular cooling of the core, it is sufficient to consider radial profiles of temperature and concentration, which are usually assumed to be isentropic and uniform, respectively. The very thin boundary layers of the outer core do not affect the bulk energy budget, contrary to those of the mantle.

The energy balance of the core equates the heat flux at the CMB to the sum of secular cooling, Q_C; latent heat from inner core crystallization, Q_L; compositional energy due to chemical separation of the inner core (often called gravitational energy, but see Braginsky and Roberts, 1995), E_ξ; and, possibly, radiogenic heat generation, Q_H. Secular cooling makes the inner core grow, which releases latent heat and compositional energy, and the first three energy sources in the balance can be related to the size of the inner core and its growth rate (Braginsky and Roberts, 1995). The current growth rate of the inner core is small (about 300 m $\mathrm{Gy^{-1}}$) and cannot be determined by observation. Thus, one has to resort to indirect means. Energy requirements for the geodynamo do not appear directly in the energy balance for the core because they are accounted for by internal energy transfers, like viscous dissipation in the mantle. The entropy balance, however, depends explicitly on dissipation (Φ_c), which results mostly from Joule heating (ohmic dissipation). Combining the

Table 11 Present state of geoneutrino observations and estimates of the total radioactivity of the Earth

	Geoneutrino flux ($10^6\,cm^{-2}\,s^{-1}$)	*Total TNU* [a]	*Crust TNU*	*Mantle TNU*	*Heat production of U and Th (total) TW*	*Reference*
Kamioka	3.4 ± 0.8	30 ± 7	23	7	11.2	Kamland collaboration (Gando, A. and 45 collaborators), 2013)
Borexino	4.4 ± 1.4	39 ± 12	23.5	15.5	15	Borexino Collaboration Group (Bellini and 88 Collaborators) (2013)

The estimated total heat generation includes the crustal contribution but not that of K.

[a]The terrestrial neutrino unit (TNU) corresponds to a flux triggering 1 event per 10^{32} protons for an exposure time of 1 year. For Th/U = 3.9, the corresponding neutrino flux is $0.113 \times 10^6\,\mathrm{cm^{-2}\,s^{-1}}$.

energy and the entropy balances, an efficiency equation can be written, which is to leading order (Braginsky and Roberts, 1995; Labrosse, 2003; Lister, 2003):

$$\Phi_c + T_\Phi \Delta S_{cond} = \frac{T_\Phi}{T_{CMB}}\left(1-\frac{T_{CMB}}{T_{ICB}}\right)Q_L + \frac{T_\Phi}{T_{CMB}}\left(1-\frac{T_{CMB}}{T_C}\right)Q_C + \frac{T_\Phi}{T_{CMB}}\left(1-\frac{T_{CMB}}{T_H}\right)Q_H + \frac{T_\Phi}{T_{CMB}}E_\xi \quad [46]$$

where T_i is the temperature at which heat due to process (i) is released and where

$$\Delta S_{cond} \equiv \int \lambda \left(\frac{\nabla T}{T}\right)^2 dV \quad [47]$$

is the entropy production due to heat conduction (see Ricard, **Chapter 7.02**).

Equation [46] shows that the different energy sources on the right-hand side serve to maintain the ohmic and viscous dissipation (Φ_c) produced at temperature T_Φ as well as the dissipation due to conduction along the isentropic temperature profile ($T_\Phi \Delta S_{cond}$). The latter term is proportional to thermal conductivity, a poorly known physical property at the (P, T) conditions of the core. Until recently, values in the range 20–77 W m^{-1} K^{-1} were used (Gubbins et al., 1982; Labrosse et al., 1997; Lister and Buffett, 1998). Recent measurements and ab initio calculations of electrical conductivity indicate larger values of thermal conductivity in the 84–140 W m^{-1} K^{-1} range at CMB conditions and in the 130–220 W m^{-1} K^{-1} range at the inner core boundary (de Koker et al., 2012; Gomi et al., 2013; Hirose et al., 2011; Pozzo et al., 2012). The rather large spread of values is not due to differences in experimental methods or theoretical calculations but to the uncertainty on core composition. The new conductivity values lead to a lower bound of 9 TW for the heat flux conducted along the isentrope at the CMB. This may be higher than the heat flux across the CMB. Were this to be true, a stable thermal stratification would develop in the upper part of the core and would impede convection and dynamo action.

All the source terms on the right-hand side of the efficiency eqn [46], save for radiogenic heating, are linked to inner core growth and are proportional to its growth rate. Therefore, if ohmic dissipation Φ_c and radiogenic heat production Q_H can be estimated, one can calculate the inner core growth rate and the heat flux across the CMB.

Ohmic dissipation in the core is dominated by the small-scale components of the magnetic field, which cannot be determined directly because they are screened by crustal magnetic sources (Hulot et al., 1997, 2002). Estimates of this term have therefore relied on dynamo models and scaling relationships. Christensen (2010) had obtained a range of $\Phi_c = 1.1 - 15$ TW (see **Chapter 8.02** by Nimmo, for a more extensive discussion). The uncertainty of this estimate affects the CMB heat flow calculation (**Figure 11**). Taking $\Phi_c = 0$ TW clearly provides the absolute lower bound and implies $Q_{CMB} = 5$ TW, which is lower than the isentropic heat flow of 9 TW. Using their lowest thermal conductivity value for the core, Gomi et al. (2013) found that $T_\Phi \Delta S_{cond} = 1.7$ TW, which is higher than the previous estimates.

The new determinations of core conductivity lead to another argument on the minimum CMB heat flux. As discussed earlier, if the conductive heat flux along the core isentrope is larger than what can be transported by sluggish mantle convection, one expects a stable thermal stratification to develop at the top of the core. This argument must be revised in the presence of compositional convection, however. Including changes of compositional energy due to light element redistribution in the energy balance for any subshell in the core, Gomi et al. (2013) had determined the vertical profile of convective heat flux required to maintain an isentropic temperature profile. Because thermal conductivity increases along a core isentrope, the conductive heat flux along an isentrope reaches a maximum at a depth of about 1000 km below the CMB. With the CMB heat flux at the isentropic value, the downward increase of the conductive heat flux must be balanced by downward convective heat flux in a thick shell at the top of the core. At this point, one can tentatively propose that the actual temperature profile in this layer deviates from an isentrope and becomes stably stratified. Some degree of stable stratification has indeed been proposed at the top of the core but over 300 km at most (see Hirose et al., 2013, for a review). In conclusion, the conductive isentropic heat flux at the top of

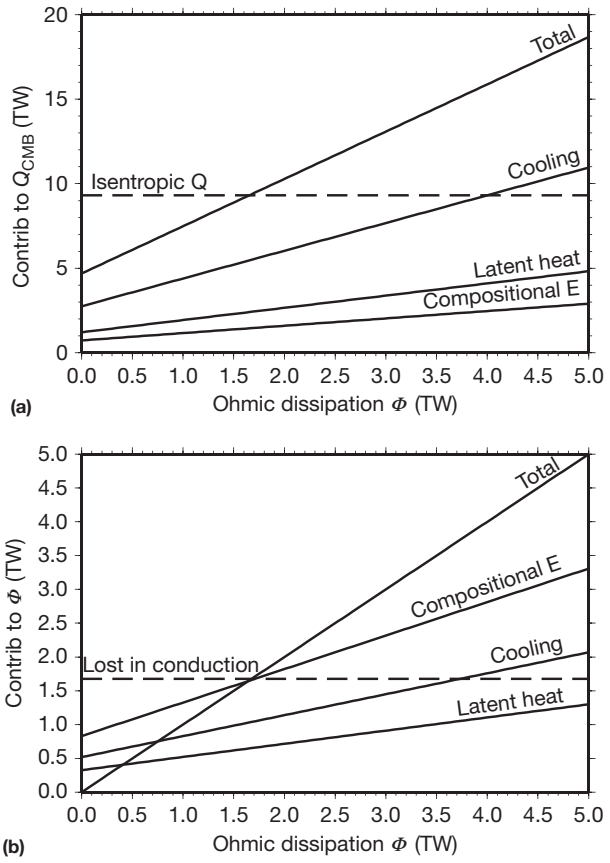


Figure 11 Contributions to the energy (a) and entropy (b) balance in the core, in the absence of internal heat production. Based on the model of Labrosse (2003) and modified by Gomi et al. (2013) to account for the new estimates of core conductivity.

the core is at least 9 TW and depends on the imperfectly known core composition; this value must be considered a lower bound for the CMB heat flow. Assuming that the CMB heat flux has been maintained at the isentropic value through geologic time, Gomi et al. (2013) had calculated that the core has cooled by at least 750 K in 4.5 Gy, a result with powerful implications that are discussed later.

7.06.5.4 Other Sources: Tidal Heating and Crust–Mantle Differentiation

The Earth's rotation is accelerating because of postglacial readjustments, and it is slowing down because of tidal interaction with the moon. The torque exerted on the Moon is due to the lag between the tidal potential and the tidal bulge, which is 2.9° ahead of the potential. In the Earth–Moon system, angular momentum is conserved, but there is a net loss of the rotational and gravitational potential energy. This energy is converted into heat by frictional forces. With laser ranging, the changes in the Earth–Moon distance have been measured accurately (3.7 cm $year^{-1}$), and the slowing down of the Earth's rotation due to tidal interaction with the Moon can be calculated exactly (Bender et al., 1973). The effect of the solar tidal potential on Earth rotation is ≈20% that of the Moon and it must be included in the calculations. The slowing down of the Earth's rotation is 5.4×10^{-22} rad s^{-1} leading to 0.024 ms $year^{-1}$ increase in the length of the day. The energy loss has been calculated to be 3 TW, which must be accounted for by dissipation in the oceans, in the solid Earth, and in the Moon. It is commonly assumed that most of the tidal friction comes from dissipation in shallow seas because dissipation in the deep oceans was shown to be small (e.g., Jeffreys, 1962; Munk and MacDonald, 1960). Lambeck (1977) calculated that dissipation in the seas and oceans must account for 90–95% of the energy dissipation. The contribution of the solid Earth tide depends on the quality factor Q and accounts for <0.1 TW for the values of Q in the mantle suggested by seismology (Zschau, 1986). Such a low value has now found confirmation from satellite observations of the lag between the solid Earth tide (0.16°) and the lunar potential. This observation implies that the dissipation by the solid Earth is 0.083 TW (Ray et al., 1996 see also **Chapter 3.09** by Gross).

Some gravitational potential energy is released by the extraction of continental crust out of the mantle. A rough estimate can be obtained as follows:

$$\delta E_g = \int_c^R g(r) r \delta\rho 4\pi r^2 dr \qquad [48]$$

where c and R are the radii of the core and Earth, respectively, and $\delta\rho$ is the density difference at radial distance r. Using a total mass of crust of 2.6×10^{22} kg and mass conservation, $\delta\rho/\rho$ is ≈0.1 in a well-mixed mantle. Neglecting induced changes of the gravity field, we get $\delta E_g \approx 2\times10^{28}$ J. For a constant rate of crustal growth during 3 Gy, the contribution to the energy budget is 0.2 TW, which can be considered as negligible. If the net extraction of crust from the mantle has occurred in a few short episodes, it has induced larger energy pulses. For three or four such episodes of ≈200−300 My duration (Condie, 1998), each pulse may have been as large as 0.6 TW, but this can still be neglected in an analysis of secular cooling. In contrast to the change of gravitational energy due to thermal contraction, most of this energy is converted to heat by viscous dissipation because it involves differential motions between melt and matrix.

7.06.5.5 Summary

Various models for the bulk silicate Earth composition lead to results that differ significantly from one another. Uncertainties on the average uranium, thorium, and potassium abundances stem from the propagation of errors through a sequence of elemental ratios and from the correction procedure for depletion effects due to melting. A lower bound on the bulk mantle heat production may be derived from the average MORB mantle source. A large uncertainty remains on the heat flux from the core, but the recently revised core conductivity values impose a higher lower bound than previously thought.

7.06.6 Secular Cooling: Constraints on Mantle Temperatures

In this section, we review evidence for secular cooling, starting with present-day evidence and working backward in time. The total heat lost by the mantle is more than all the inputs. Our preferred values for the input and output of energy are 23 and 39 TW, respectively (**Table 12**). The difference 16 TW must be accounted for by the secular cooling of the mantle (**Figure 12**). Addition of all the uncertainties results in an unrealistically wide range (1–28 TW) for the secular cooling. Assuming a constant value for the specific heat of 1250 J kg^{-1} K^{-1}, the rate of cooling must be 3.4×10^{-15} K s^{-1} or 106 K Gy^{-1}, with a range 7–190 K Gy^{-1}.

Table 12 Mantle energy budget, preferred value and range

	TW	*TW*
Oceanic heat loss (300×10^6 km^2)	32	30–34
Continental heat loss (210×10^6 km^2)	14	13–15
Total surface heat loss (510×10^6 km^2)	46	43–49
Radioactive sources (mantle + crust)	18	13–23
Continental heat production (crust + lith. mantle)	8	7–8
Heat flux from convecting mantle	38	35–41
Radioactive heat sources (convecting mantle)	11	9–17
Heat from core	11	5–17[a]
Tidal dissipation in solid Earth	0.1	
Gravitational energy (differentiation of crust)	0.3	
Total input	23	14–34
Net loss (mantle cooling)	16	1–29
Present cooling rate, K Gy^{-1}	106	7–210
Present Urey ratio[b]	0.29	0.12–0.49

The distribution in the range is barely known for most cases and the preferred value is simply the middle one. The cooling rate is computed assuming $C_P = 1200$ J K^{-1} kg^{-1}.
[a]This range includes estimates from core thermodynamics and inference from the perovskite–postperovskite phase diagram.
[b]Urey ratio for the convecting mantle, leaving out crustal heat sources from both the heat loss and the heat production.

7.06.6.1 The Present-Day Mantle Geotherm

The potential temperature of shallow oceanic mantle may be calculated from the composition of MORBs which have not been affected by fractional crystallization, and also from heat flux and bathymetry data, as explained earlier. Such independent determinations are summarized in **Table 4** and are in very good agreement with one another. For temperatures at greater depth, one may use seismic discontinuities and the associated solid-state phase changes. This classical method requires specification of the mantle composition, which is usually taken to be pyrolite (Ringwood, 1962). Well-defined discontinuities at depths of 410 and 660 km have been linked to the olivine–wadsleyite transition and to the dissociation of spinel to ferropericlase and magnesiowüstite, the so-called postspinel transition. Other seismic discontinuities have been identified, notably at a depth of about 500 km (**Table 13**). Some of these are not detected everywhere and seem to have a regional character, and interpretation is still tentative to some extent. Recently, a new phase change relevant to the lowermost mantle has been discovered, from perovskite (Pv) to postperovskite (pPv) (Murakami et al., 2004; Oganov and Ono, 2004; see also **Chapter 2.08** by Oganov and **Chapter 2.05** by Yuen et al.).

The olivine–wadsleyite phase change has a large Clapeyron slope in a 3–4 MPa K^{-1} range and hence provides an accurate temperature estimate (Katsura et al., 2004). Errors on temperature arise from the experimental data and from small differences in the depth of the seismic discontinuity. In addition, in a multicomponent material such as the mantle, there exists a transition region whose thickness depends on composition (Stixrude, 1997). Accounting for these effects, the temperature estimate at 410 km is 1760 ± 45 K for a pyrolytic upper mantle. Isentropic profiles that pass through these (P, T) values (**Figure 13**) correspond to potential temperatures in the range 1550–1650 K (Katsura et al., 2004), in very good agreement with the independent estimates from the composition of MORBs and heat flux data (**Table 4**).

Laboratory studies have cast doubt on the postspinel transition pressure (Irifune and Isshiki, 1998; Katsura et al., 2003). Furthermore, this transition may have a very small Clapeyron slope (as small as -0.4 MPa K^{-1} according to Katsura et al. (2003)), implying that small uncertainties on pressure lead to large errors on the transition temperature. For these reasons, one should treat temperature estimates for this transition with caution. Nevertheless, the uncertainty often quoted for this value is similar to that for the 410 km discontinuity, and the

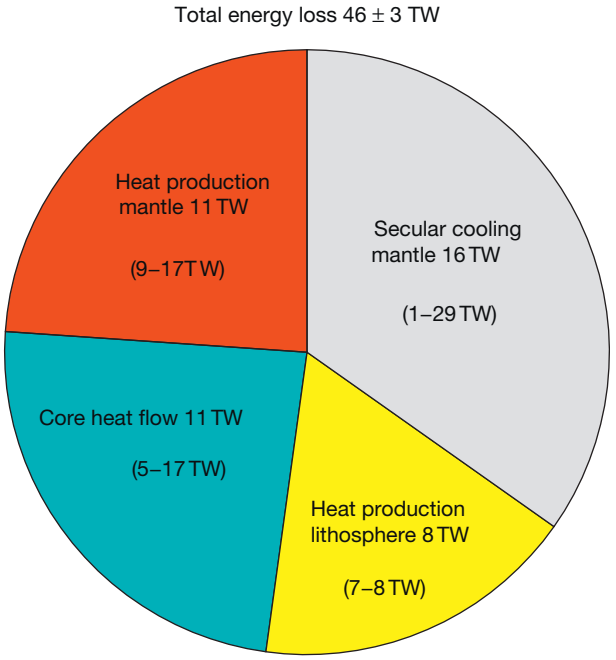


Figure 12 Proposed breakdown of the present energy budget of the Earth. With continental lithosphere, mantle heat production, and core heat loss constrained, the mantle cooling rate is adjusted to fit the total energy loss.

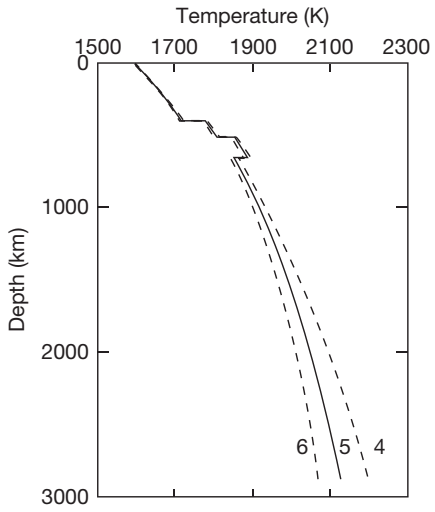


Figure 13 Isentropic temperature profiles in the mantle for different values of the Anderson–Grüneisen parameter, as labeled. Reproduced from Katsura T, et al. (2004) Olivine–Wadsleyite transition in the system $(Mg,Fe)SiO_4$. *Journal of Geophysical Research* 109: B02209.

Table 13 Anchor points for the mantle geotherm

Boundary	*Depth (km)*	*Temperature (K)*	*Reference*
MORB generation	50	1590–1750 [a]	Kinzler and Grove (1992)
Olivine–wadsleyite	410	1760 ± 45	Katsura et al. (2004)
Postspinel	660	1870 ± 50	Katsura et al. (2003, 2004)
Core–mantle	2900	4080 ± 130	Alfé et al. (2002) and Labrosse (2003), this paper

[a]True range of temperatures in the shallow mantle.

present-day 'best' value for the mantle temperature at a depth of 660 km is 1870 ± 50 K (Ito and Katsura, 1989; Katsura et al., 2003, 2004).

The other major discontinuity, the core–mantle boundary, is a chemical boundary and its temperature can be computed from the core side. Once again, the method relies on a phase change, in this case the solidification of iron at the inner core boundary. Using ab initio calculations, Alfé et al. (2002) had determined that the melting temperature of pure Fe is in the range 6200–6350 K. Experiments by Anzellini et al. (2013) up to a pressure of 200 GPa are compatible with this value. In the presence of light elements (O, Si, and S), the liquidus temperature at the inner core boundary is lowered by 700 ± 100 K. The outer core can be assumed to be very close to isentropic (e.g., Braginsky and Roberts, 1995), and the variation in temperature can be linked to the variation in density by (e.g., Poirier, 2000)

$$\left(\frac{\partial T}{\partial \rho}\right)_S = \gamma \frac{T}{\rho} \qquad [49]$$

where γ is the Grüneisen parameter (see **Chapter 7.02**). According to the theoretical calculations of Vočadlo et al. (2003; see also Vočadlo, **Chapter 2.06**), $\gamma = 1.5 \pm 0.01$ throughout the core. We may therefore assume that γ is constant and integrate eqn [49]:

$$T_{CMB} = T_{ICB}\left(\frac{\rho_{CMB}}{\rho_{ICB}}\right)^{\gamma} = 0.73 T_{ICB} \qquad [50]$$

which relates the temperature at the CMB to that at the ICB and to density structure. Using $\gamma = 1.5$ and density values from the preliminary reference Earth model (PREM) of Dziewonski and Anderson (1981; see also Dziewonski and Romanowicz, **Chapter 1.01**), we obtain a range of 3950–4220 K for the temperature at the top of the core. This does not account for uncertainties in the PREM density values, and the range is narrower than the full range of published values, but a full discussion of all the different estimates is outside the scope of this chapter.

Strong constraints on temperatures above the CMB have been deduced from studies of the Pv to pPv phase change, provided that it takes place in the mantle (Murakami et al., 2004; Oganov and Ono, 2004; see also Irifune, vol. 2). According to the experiments of Tateno et al. (2009), this phase change has a Clapeyron slope of 13.3 ± 1.0 MPa K^{-1} and a temperature of 3520 ± 70 K at the CMB pressure. A key feature is that it is likely to occur in the so-called D'' boundary layer at the base of the mantle. With a simple model for the radial temperature profile in this boundary layer, Hernlund et al. (2005; see also **Chapter 1.22** by Lay and Hernlund and McNamara, **Chapter 7.11**) predicted that the phase change boundary should be crossed twice in cold regions of the mantle (**Figure 14**) and should not be crossed in hot regions, which seems to be consistent with seismic observations. As illustrated in **Figure 14**, a double crossing of the phase boundary implies that the temperature gradient at the base of the mantle is higher than that of the Clapeyron diagram. It also implies that the temperature at the CMB must be higher than that of the phase change at that pressure (Hernlund and Labrosse, 2007). This provides us with a lower bound of 3700 K on the CMB temperature (Tateno et al., 2009), which is consistent with estimates obtained from the core side. Lateral variations of the depth of discontinuity that are determined by seismological observations imply lateral temperature differences of ≈ 1500 K in the lowermost mantle (Hirose, 2006).

Paired seismic discontinuities have also been detected in hot regions near the core–mantle boundary beneath the Pacific Ocean (Lay et al., 2006). These can be reconciled with the model of Hernlund et al. (2005) if these regions are enriched in iron. Using a model for the boundary layer above the core–mantle boundary and a thermal conductivity value of 10 Wm^{-1}K^{-1}, Lay et al. (2006) had estimated that the heat flux across the CMB is at least 13 ± 4 TW.

The interpretation of D'' discontinuities as due to the Pv to pPv phase transition has been challenged by several experimental studies. Catalli et al. (2009), Andrault et al. (2010), and Grocholski et al. (2012) had found that the Pv and pPv phases can stay in equilibrium with each other over a wide pressure range, which is not consistent with the sharpness of

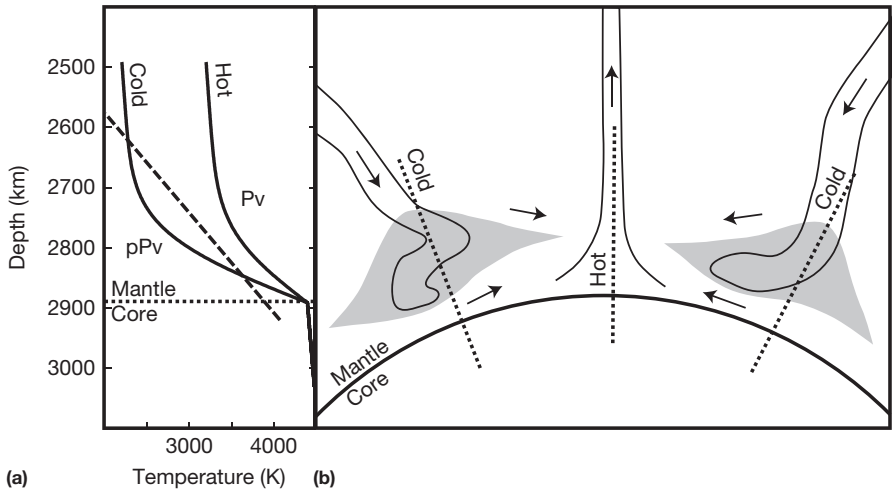


Figure 14 Sketch of postperovskite lenses in cold regions of the lowermost mantle from Hernlund et al. (2005). (a) Schematic temperature profiles for 'cold' and 'hot' regions (plain curves) superimposed onto the phase boundary. (b) Sketch of thermal boundary layer at the base of the mantle with convective downwellings and upwellings.

the seismic discontinuity. These results are in direct contradiction with those of Tateno et al. (2009) who reported a sharp phase transition. The region of Pv–pPv coexistence could extend through the whole D'' layer and, for some mantle compositions, could be found at pressures higher than that of the CMB (Grocholski et al., 2012).

7.06.6.2 Temperature Versus Time

One may use petrologic constraints to investigate past temperatures of the mantle. Continental crustal material was different in the Archean than it is today. Basaltic lavas exhibit systematic compositional trends with time, including a secular decrease in average MgO content. MgO-rich ultramafic lavas named komatiites are common in the Archean and are almost absent from today's rock record. Early workers proposed that komatiites require the mantle source to be at least 300 K hotter than present (Green, 1975; Sleep, 1979), but they considered dry mantle only. The peridotite solidus depends strongly on water content, which in turn depends on the geologic setting. If komatiites are generated by mantle plumes, involving mantle that is essentially dry, that is, such that its water content is so small that it does not affect phase boundaries, one deduces that mantle plume temperatures have decreased by about 300 K in 3 Gy (Nisbet et al., 1995). Jarvis and Campbell (1983) suggested that such hot mantle plumes did not require the Archean mantle to be more than 100 K hotter than present on average. According to an alternative hypothesis, komatiites are generated in a subduction environment, involving mantle hydrated by downgoing plates. In that case, one is led to conclude that this part of the mantle was only slightly hotter ($\approx$100 K) in the Archean than it is today (Grove and Parman, 2004). In both cases, komatiites do not sample 'average' mantle and it is not clear how to incorporate these temperature estimates in models for the entire mantle.

Mid-ocean ridge tholeiites are better suited for studies of the mantle's average temperature because they can be sampled over very large areas. They are a compositionally heterogeneous group, however, which translates into a wide temperature range ($\approx$200 K) (Kinzler and Grove, 1992; Klein and Langmuir, 1987). Abbott et al. (1994) calculated the liquidus temperature for Phanerozoic MORBs and Archean MORB-like greenstones and determined the maximum and minimum mantle potential temperatures versus time. Although the range of temperatures for each period is wide ($\approx$200 K), the trend is well marked. Abbott et al. (1994) concluded that mantle temperatures decreased by $\approx$150 K since 3 Gy, which is less than the range of mantle potential temperatures at a given time. Work by Herzberg et al. (2010) on non-arc basaltic rocks, which are again chosen because they were melts of the ambient mantle, confirms these values (**Figure 15**). A cooling rate of 50 K Gy^{-1} for the mantle represents $\approx$8 TW.

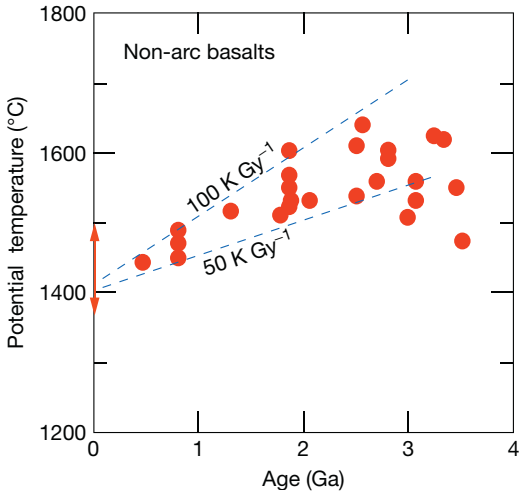


Figure 15 Temperature of the source of non-arc basaltic rocks as a function of age. The vertical arrow at zero age stands for the range of mantle potential temperature estimates from Table 4. Most of the data fall between the 50 K Gy^{-1} and the 100 K Gy^{-1} lines. Data from Herzberg et al. (2010).

7.06.6.3 Early Earth

A full description of the Earth's thermal evolution must include the initial conditions. Here, 'initial' refers to the time when the Earth had completed its main phase of core–mantle differentiation and the mantle had solidified to the point where its dynamics can be described as subsolidus convection. Before reaching that point, a host of processes with different dynamics occurred. They may be separated in three categories: accretion, core formation, and magma ocean crystallization. The process of the formation of the Earth brought together matter that was originally dispersed in the protosolar nebula, thereby releasing gravitational energy. One may estimate the total energy released by taking the difference between the total gravitational energies before and after. The fate of this energy, however, depends on the way it is dissipated and transformed into another type of energy. The effect of core differentiation is quite different from that of accretion. Most of the processes involved remain speculative to some extent, and we restrict our discussion to the points that are directly relevant to the early thermal structure. Several review articles (e.g., Stevenson, 1989; Wetherill, 1990, several articles in vol. 9), and two books (e.g., Canup and Righter, 2000; Newsom and Jones, 1990) deal with these issues in detail.

During accretion, the gravitational energy of impactors is first transformed into kinetic energy and then dissipated in the form of heat at the impact. One may define two limit cases. If no energy is lost to space or stored as elastic energy, the temperature of the whole Earth is raised by an amount equal to

$$\Delta T = \frac{-E_g}{MC_P} \sim 3.75 \times 10^5 \text{K} \qquad [51]$$

which would be sufficient to vaporize the whole planet. A large fraction of the impact energy, however, is released at shallow levels and lost to space by radiation. Stevenson (1989) estimated that, if all the energy is made available for radiation, accretion would raise the temperature of the Earth by $<$70 K relative to that of the nebula. The actual evolution lies somewhere between these two limiting cases, involving partial dissipation of the impact energy within the planet and radiative heat transfer through the primordial atmosphere. One important

factor is the size of the impactors. 'Small' impactors, which are much smaller than the target, account for the vast majority of impacts on Earth after the planetary embryo stage (Melosh and Ivanov, 1999). The depth of energy release increases with the size of the impactor and one key variable is the ratio between the time for energy transport to the surface and the time between two impacts. During accretion, evolution toward larger and fewer impactors has two competing effects: energy gets buried at greater depth while the time between two impacts increases, which enhances heat loss to the atmosphere. Assuming heat transport by diffusion, Stevenson (1989) concluded that typical accretion scenarios lead to significant energy retention within the planet. The extreme case is that of the giant impact thought to be at the origin of the Moon. Calculations suggest that the whole Earth temperature was raised to as high as 7000 K (Cameron, 2001; Canup, 2004). In such conditions, the whole Earth melted and parts of it were vaporized to form a thick atmosphere. The question of whether or not previous impacts were able to melt the Earth becomes irrelevant.

The formation of the core also has important energetic implications. Some gravitational energy is released by going from a homogeneous material to a stratified core–mantle system. Kinetic energy plays no role in this process, in contrast to the accretionary sequence, and gravitational potential energy is dissipated by viscous heating in both the iron and the silicate phases. Flasar and Birch (1973) estimated that this process would heat the whole Earth by about 1700 K. This estimate relies on the bulk difference in gravitational energy between the initial and the final states and hence gives no information on where energy gets dissipated. We know now that iron–silicate differentiation occurred very early in the solar system and affected planetesimals (Kleine et al., 2002). Clearly, core formation within planetesimals and in the Earth after the giant impact involves different dynamics. In large-scale models of the giant impact, the cores of the two protoplanets merge with little mixing with silicates (Canup, 2004). An iron emulsion may form in the molten silicate, however, due to a combination of the Rayleigh–Taylor and Kelvin–Helmholtz instabilities (Ichikawa et al., 2010; Rubie et al., 2003).

Several mechanisms have been proposed for core formation (e.g., Stevenson, 1990; Rushmer et al., 2000; see also Rubie **Chapter 9.03**): percolation of liquid metal through a compacting solid silicate matrix, iron droplets 'raining' through a magma ocean (Ichikawa et al., 2010), diapirs generated by Rayleigh–Taylor instability at a rheological interface, and hydraulic fracturing (Stevenson, 2003). All these mechanisms may have been active simultaneously or in succession at different times (Golabek et al., 2008; Ricard et al., 2009; Šrámek et al., 2010). They have different implications for dissipation. For example, percolation and compaction can explain core formation in planetesimals and planetary embryos (Šrámek et al., 2012). On planets larger than about the size of Mars, the same process can operate locally below impact sites and generate a metallic diapir, which sinks into the deeper mantle once enough metal has accumulated. Gravitational energy release during descent leads to further melting and can trigger runaway core growth (Monteux et al., 2009; Ricard et al., 2009).

For a Newtonian rheology, the amount of viscous heating is $\phi \sim \eta U^2/L^2$, where η is the viscosity of the fluid phase, which may be silicate or iron, and U and L are the scales for velocity and length. In the case of an iron diapir, the velocity and length scales are the same for the metal and silicate phases, but the viscosity of the former is several orders of magnitudes lower. Viscous heating is thus concentrated in the silicate phase and produces little heating of the iron phase because the heat diffusion timescale is less than the descent timescale. This would differentiate a core that is initially colder than the lower mantle. In the case of interstitial flow, the small size of iron veins makes heat diffusion very effective and thermal equilibration with the surrounding silicate phase likely. In the case of iron droplets raining down through molten silicate, the droplet size is set by a balance between surface tension and viscous drag and is typically 1 cm (Ichikawa et al., 2010; Ulvrová et al., 2011). Thus, one also expects thermal equilibrium between metal and silicate. With these two last mechanisms, the metallic core is generated at the temperature of the lower mantle. In the model of Ricard et al. (2009), the diapir is in fact a two-phase mixture of liquid metal and solid silicate that remain in thermal equilibrium. This type of diapir heats up the surrounding solid silicate during descent. The resulting thermal structure is that of a hot core and lower mantle, a cool midmantle, and a hot upper mantle. Thus, the Earth's initial thermal structure depends strongly on the mechanism of core formation and cannot be specified with certainty yet.

7.06.6.4 Magma Ocean Evolution

Both the giant impact and core formation generated temperatures that were high enough for the entire silicate Earth to be molten. Cooling and crystallization of such a thick magma ocean involve heat transfer through the primordial atmosphere, convection, rotation, and crystal–melt separation. Available models have been aimed mostly at determining the extent of chemical stratification at the end of crystallization (Abe, 1997; Solomatov, 2000).

At the scale of the whole Earth's mantle, solidification and crystal–melt separation involve pressure-dependent properties that vary by large amounts from top to bottom. Current knowledge on a few of these is not sufficient to draw firm conclusions on magma ocean evolution. Starting from a fully molten mantle where crystallization begins, for example, depends on the respective slopes of the liquidus and the isentrope. We know that the liquidus is steeper than the isentrope over about 1000 km at the top of the mantle, but the reverse may be true at greater depths (Andrault et al., 2011; Fiquet et al., 2010; Thomas et al., 2012). An equally important variable is the density difference between melt and solid at equilibrium, which controls the manner of solid–liquid segregation. The standard magma ocean scenario postulates that crystallization proceeds from the bottom upward, which requires that crystals sink through melts at all depths. Melts are less dense than solids at low pressures, but the large compressibility of liquids compared to that of crystals implies that the latter could be buoyant with respect to silicate melts in the lower mantle (Andrault et al., 2012; de Koker et al., 2013; Nomura et al., 2011; Thomas et al., 2012). In this case, there might have been two different magma oceans in the early Earth at the top and bottom of the mantle, separated by a growing solid midmantle horizon (Labrosse et al., 2007).

The low viscosity of the melt and the size of Earth imply highly turbulent convective flows and rapid cooling, such that

the magma ocean at the surface solidifies in a few thousand years (Abe, 1997). A basal magma ocean, on the other hand, would cool at a much slower rate due to sluggish heat transfer across the solid mantle and could still contain isolated pockets of partial melt today (Labrosse et al., 2007). This would account for the ultralow-velocity zones that have been found above the CMB (Williams and Garnero, 1996).

Turning to the evolution of a surficial magma ocean, one must account for two rheological transitions, from pure magma to slurry (crystals suspended in a magma) and from slurry to mush (interconnected crystals forming a compacting matrix), which affect the convection regime and the cooling rate. Starting from a superheated melt at temperatures above the liquidus, the initial phase has a solid growing at the base and a fully molten upper layer that becomes thinner as cooling proceeds. A first transition occurs when the fully molten layer vanishes. At this stage, the Earth is made of a partially crystallized magma ocean that may lie over already fully solidified mantle. The radial temperature profile is tied to the solidus that has a steeper slope than the isentrope. Such a thermal stratification is unstable and leads to the convective overturn of the solid layer (Elkins-Tanton et al., 2003). The two layers evolve at vastly different timescales because of their different rheologies. The bulk cooling rate is set by the heat loss through the Earth's surface, which is controlled by the dynamics of the partially crystallized magma ocean at the top. In this second phase, heat transport occurs mostly by melt–solid separation and solidification proceeds from the bottom up. The fully solidified layer at the base of the magma ocean thickens rapidly and eventually becomes unstable. Convective overturn is slower than the cooling of the magma ocean and may be considered as a separate event that leads to decompression melting and the formation of a secondary magma ocean at the surface and possibly at the bottom of the mantle. The process of cooling and solidification of this magma ocean then repeats itself. This regime prevails until the shallow magma ocean reaches the rheological threshold between liquid behavior and solid behavior, which probably occurs at a crystal fraction of about 60%. At this stage, the shallow partially crystallized layer becomes strongly coupled to the solid mantle below and cooling proceeds through bulk convection everywhere. According to Abe (1993, 1997), this was completed in a few 10 My (**Figure 16**) and sets the initial conditions for secular cooling models of the solid Earth. From the most recent phase diagrams (Herzberg and Zhang, 1996; Litasov and Ohtani, 2002), the final rheological transition corresponds to a potential temperature of about 1800 ± 100 K for a mantle composed of dry pyrolite (**Figure 17**).

Recent advances in geochemistry have confirmed the theoretical estimates of Abe (1997). Caro et al. (2003) had found evidence that pushes early crust formation as far back as 4.4 Gy. Others have argued that liquid water was present on the Earth's surface at 4.3 Gy (Mojzsis et al., 2001). These studies cannot demonstrate that plate tectonics was already active at such early times, but provide some support for a solid upper mantle.

7.06.6.5 Average Secular Cooling Rate of the Mantle

Subsolidus convection began at a mantle potential temperature of about 1800 ± 100 K, which exceeds the present-day temperature by about 200 K. This constrains the average cooling rate of the Earth even if the timing is not known precisely. These various constraints lead to an average cooling rate of about 50 K Gy^{-1} that is consistent with changes in the compositions of erupted lavas (Abbott et al., 1994; Herzberg et al., 2010).

The present-day mantle potential temperature is fixed at about 1600 K by a fit to heat flux and bathymetry data independently of petrologic constraints and in particular of the

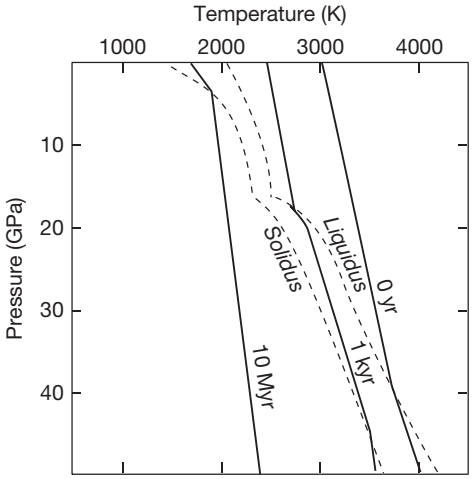


Figure 16 Three geotherms at different times in the early Earth. The timescale for the thermal evolution is set by heat loss at the upper boundary that decays rapidly with temperature. At about 10 My, the solid content in the partially molten upper mantle layer reaches the threshold value of 60%, which marks the cessation of liquid behavior. After that time, convection is in the subsolidus regime controlled by solid behavior that still prevails today. Adapted from Abe Y (1997) Thermal and chemical evolution of the terrestrial magma ocean. *Physics of the Earth and Planetary Interiors* 100: 27–39.

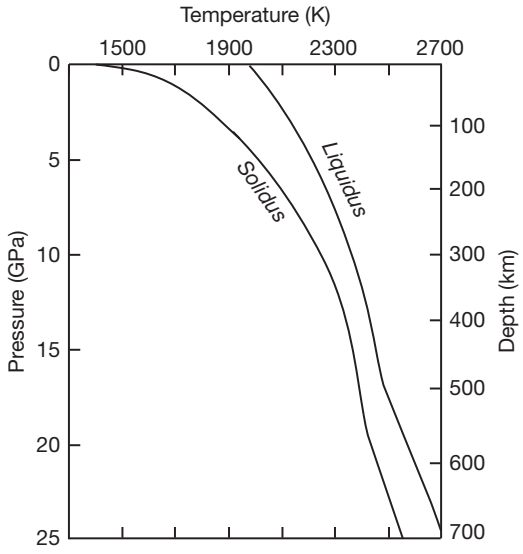


Figure 17 Solidus and liquidus for dry pyrolite as a function of pressure. Adapted from Litasov K and Ohtani E (2002) Phase relations and melt compositions in CMAS–pyrolite–H_2O system up to 25 GPa. *Physics of the Earth and Planetary Interiors* 134: 105–127.

water content of mantle rocks (McKenzie et al., 2005). If the mantle contained significant amounts of water at the end of the magma ocean phase, however, the phase diagram must be shifted to lower temperatures. One consequence is that the potential temperature at the beginning of subsolidus convection was lower than the 1800 K estimate used earlier. In this case, the average cooling rate must be even $< 50\ \mathrm{K\ Gy^{-1}}$.

7.06.6.6 Summary

At the end of accretion and core–mantle separation, a magma ocean probably extended through a large fraction of the silicate Earth. Because of its small viscosity, it cooled and crystallized rapidly, which led to a stratified mantle with a solid lower layer and a partially crystallized upper layer. At the start of solid-state mantle convection, upper mantle temperatures were such that the surficial partially molten region had a solid fraction of about 60%. This sets the initial temperature of the solid Earth to a value that is about 200 K higher than the present.

7.06.7 Thermal Evolution Models

7.06.7.1 The Urey Ratio

The decay time of the bulk radiogenic heat production, which is the weighted average of the individual decay times of the four relevant isotopes (Table 9), is 3 Gy. Thus, over the Earth's history, heat sources have decreased by a factor of about 4. The efficiency of the Earth's convective engine in evacuating heat generated by radioactive decay is commonly measured by the Urey ratio, *Ur*, which is the ratio of heat production over heat loss:

$$Ur = \frac{\int_V H \mathrm{d}V}{-\int_A \mathbf{q}\cdot\mathbf{n}\mathrm{d}A} \qquad [52]$$

To calculate this ratio, we do not take continental heat sources into account because they are stored in the continental lithosphere and hence are not involved in mantle convection. Using the data of Table 12, we find that $Ur = 0.29$, with a range of 0.12–0.49.

The heat budget of Table 12 allows calculation of the present-day cooling rate. Secular variations of basalt compositions and consideration of initial thermal conditions provide constraints on the total temperature drop over the Earth's history and hence on the average cooling rate (Figure 15). Thus, physical models are not needed to determine how the Earth has cooled down. Instead, the data allow a test of our understanding of mantle convection processes, and available constraints on the cooling rate can be turned into one for how the rate of heat loss has changed through time. The global heat balance reads as

$$M\langle C_\mathrm{p}\rangle \frac{\mathrm{d}T}{\mathrm{d}t} = -Q + H \qquad [53]$$

where M is the mass of the Earth and $\langle C_\mathrm{p}\rangle$ an 'effective' heat capacity that accounts for the isentropic variation of temperature with depth. Integrating over the age of the Earth, one deduces that

$$\frac{\overline{Q} - \overline{H}}{Q - H} = \frac{(\mathrm{d}T/\mathrm{d}t)_\mathrm{av}}{\mathrm{d}T/\mathrm{d}t} \qquad [54]$$

where $\overline{Q}$ and $\overline{H}$ are the time-averaged values of heat loss and heat production and $(\mathrm{d}T/\mathrm{d}t)_\mathrm{av}$ is the average cooling rate. The cooling rate has an average value of about $50\ \mathrm{K\ Gy^{-1}}$ (i.e., about 200 K over 4 Gy; see Section 7.06.6.6) and a larger present-day value (about $120\ \mathrm{K\ Gy^{-1}}$). Thus, the ratio in eqn [54] is < 1 and probably as low as 0.4. This implies that the rate of heat loss has decreased less rapidly than that of heat production.

7.06.7.2 'Parameterized' Cooling Models

The characteristics of convection in well-mixed homogeneous layers are well understood and have been used to develop secular cooling models for Earth. Here, we briefly recapitulate a few important results and set up the stage for further discussion with simple arguments. With constant physical properties, the heat flux through the top boundary of a convecting layer is a function of the temperature difference across that layer (e.g., Davies, 1980a,b; Schubert and Young, 1976; Sharpe and Peltier, 1978). A very robust scaling law relates the dimensionless heat flux (i.e., the Nusselt number) to the Rayleigh number:

$$Nu = \frac{Q/A}{kT/D} = C_1 Ra^{\beta} \qquad [55]$$

where C_1 is a proportionality constant, Q/A the heat flux, A the surface area, T the temperature difference across the layer, D the layer thickness, and Ra the Rayleigh number:

$$Ra = \frac{g\alpha T D^3}{\kappa \nu_\mathrm{M}} \qquad [56]$$

where $\nu_\mathrm{M} = \mu_\mathrm{M}/\rho$ is the kinematic viscosity. This relationship can be turned into an equation for heat loss Q of the form

$$Q = C_2 T^{1+\beta} \nu_\mathrm{M}^{-\beta} \qquad [57]$$

where the constants C_2 and β are obtained from boundary layer theory as well as laboratory experiments (Howard, 1964; Olson, 1987; Turcotte and Oxburgh, 1967). This relationship is valid if, and only if, instability always occurs in the same conditions, for example, when a Rayleigh number defined locally in the boundary layer exceeds a critical value. Cooling models of this kind have been termed 'parameterized' because they collapse all the physics of mantle convection into a single equation involving only temperature and two dimensionless parameters, C_2 and β. Typically, $\beta = 1/3$, such that heat loss is governed solely by local instabilities of the upper boundary layer. The value of constant C_2, but not that of exponent β, depends on the instability threshold and on the length of the convective cell (Grigné et al., 2005; Olson, 1987).

One additional parameter is required to account for the strong dependence of viscosity on temperature. Secular cooling implies that the mantle viscosity increases with time, which slows down convection and decreases the cooling rate. One can approximate an Arrhenius law for viscosity by an equation of the form $\nu = \nu_0 (T/T_0)^{-n}$, with $n \sim 35$, which is valid for $T \sim T_0$ (Christensen, 1985; Davies, 1980a,b). The thermal evolution equation then takes the following form:

$$M\langle C_p\rangle\frac{dT}{dt}=-Q_0\left(\frac{T}{T_0}\right)^{1+\beta(1+n)}+H(t) \quad [58]$$

where Q_0 is the heat loss at the reference potential temperature T_0. Temperature changes in the Earth are small compared to the absolute temperature (i.e., $\approx$200 K for a present-day temperature of $\approx$1600 K). One may thus linearize the thermal evolution equation:

$$T=T_0+\Theta, \text{ with } \Theta \ll T_0 \quad [59]$$

$$M\langle C_p\rangle\frac{d\Theta}{dt}=-Q_0\left[1+\frac{\Theta}{T_0}(1+\beta+\beta n)\right]+H(t) \quad [60]$$

For simplicity, we assume that heat production decreases exponentially with time, such that $H(t)=H_0\times\exp(-t/\tau_r)$, where $\tau_r\approx 3$ Gy. The solution of eqn [60] is

$$\Theta=\Theta_0\times\exp(-t/\tau_p)+\frac{Q_0\tau_p}{M\langle C_p\rangle}(\exp(-t/\tau_p)-1)+\frac{H_0\tau_p\tau_r}{M\langle C_p\rangle T_0(\tau_r-\tau_p)}(\exp(-t/\tau_r)-\exp(-t/\tau_p)) \quad [61]$$

where we introduce the convective relaxation time τ_p:

$$\tau_p=\frac{M\langle C_p\rangle T_0}{(1+\beta+\beta n)Q_0} \quad [62]$$

Using standard values for the parameters and variables involved, $n=35$, $\beta=1/3$, $M=6\times10^{24}$ kg, $Q_0=30$ TW, $T_0=1300$ K (the temperature jump across the boundary layer is the relevant parameter here), and $\langle C_p\rangle=1200\,\mathrm{J\,kg^{-1}\,K^{-1}}$, the thermal relaxation time is about 800 My.

Equation [61] for the average mantle temperature depends on a number of parameters. We can summarize simply the constraints brought by observations by noting that, for $t\gg\tau_p$,

$$Ur\rightarrow\frac{(\tau_r-\tau_p)}{\tau_r} \quad [63]$$

For $\tau_p=800$ My and $\tau_r=3$ Gy, $Ur=0.75$, which is much larger than the values derived earlier. In order to meet the constraint of the Urey ratio, one must increase the adjustment time of mantle convection. Rephrased differently, this states that the rate of heat loss does not keep up with the decrease of radiogenic heat production. We have reached the same conclusion in the previous section with a completely different argument on the difference between the secular average and the present-day values of the cooling rate.

In order to increase the adjustment time of mantle convection, one option is to appeal to a layered mantle (McKenzie and Richter, 1981). Another option is to decrease the value of exponent β. According to Christensen (1984, 1985), this may be a consequence of the temperature dependence of viscosity. For very large variations of viscosity, however, convection occurs in the stagnant lid regime (e.g., Davaille and Jaupart, 1993; Ogawa et al., 1991; Solomatov and Moresi, 1997). In this case, plate tectonics is shut off, which is not a satisfactory solution. There may be an intermediate regime with subduction of the very viscous lid, such that $\beta=0.293$ (Solomatov and Moresi, 1997), but this β value is too large to meet the constraint of the Urey ratio. In an attempt to account for the resistance to plate bending at subduction zones, Conrad and Hager (1999) had suggested that $\beta\sim0$. Following a similar line of reasoning and building on a study by Sleep (2000), Korenaga (2003) pushed the argument further by introducing variations of plate rigidity due to changes in the depth and degree of melting as the mantle cools down. He proposed that β is in fact negative. In accurate models of compressible mantle convection, however, dissipation due to plate bending accounts for $<10\%$ of the total dissipation and has little impact on the value of β (Leng and Zhong, 2010). Furthermore, there is no evidence that resistance to bending actually limits subduction on Earth. For example, no correlation exist between the slab dip (which specifies the radius of curvature of the plate at the trench and hence the extent of bending) and the magnitude of slab pull or the convergence rate (Lallemand et al., 2005).

'Parameterized' models rely on strong assumptions on plate size and/or subduction age, which are difficult to test and which may not be valid. One important point is that, after about 2 Gy, temperatures from eqn [61] are not sensitive to the initial conditions, which has two implications. One is that failure to reproduce the present-day Urey ratio cannot be blamed on the poorly known initial condition. The other implication is that 'backward' thermal calculations starting from the present become unreliable for old ages.

We now describe several features of the Earth's convective system that make it quite distinctive. We discuss the relevance of heat loss 'parameterizations' and evaluate whether the present-day heat balance is representative of the secular evolution of Earth, which, by definition, only involves long-term changes.

7.06.7.3 The Peculiarities of Mantle Convection: Observations

Convection in the Earth's mantle proceeds in peculiar ways. One distinctive feature is the triangular age distribution of the seafloor (**Figure 7**; Becker et al., 2009; Labrosse and Jaupart, 2007; Parsons, 1982), which is at odds with other convecting systems as well as with the parameterized schemes discussed earlier. We show in **Appendix H** the age distribution at the upper boundary of several convective layers. None of them resemble that of Earth, which illustrates current limitations in reproducing mantle convection processes.

A few other peculiar features of mantle convection are worth mentioning. Heat loss is unevenly distributed at the surface. The Pacific Ocean alone accounts for almost 50% of the oceanic total and 34% of the global heat loss of the planet. This is due in part to the large area of this ocean and in part to its high spreading rate. Oceanic plates are transient, such that changes of oceanic heat loss may occur when a new ridge appears or when one gets subducted. For example, the heat flux out of the Atlantic Ocean is about 6 TW, 17% of the oceanic total (Sclater et al., 1980). This ocean has almost no subduction and started opening only at 180 My. At that time, the generation of a new mid-ocean ridge led to an increase of the area of young seafloor at the expense of old seafloor from the other oceans and hence to enhanced heat loss. From the standpoint of dynamics, the most challenging features of mantle convection are perhaps the large variations of plate speeds and dimensions that exist. With the small number of

plates present, averaging values of spreading velocity and plate size may well be meaningless.

One key feature of Earth is the presence of large continents at the surface, which play an important role. They do not allow large heat fluxes through them and generate boundaries with complicated shapes that constrain mantle flow. They exert a strong control on secular cooling and may well be responsible for the triangular distribution of ocean floor ages (Coltice et al., 2012; Grigné and Labrosse, 2001; Labrosse and Jaupart, 2007; Lenardic et al., 2005).

7.06.7.4 Convection with Oceanic Plates

Controls on the Earth's heat loss are direct consequences of the convection regime. To derive an equation for heat loss on Earth, we focus on the oceans. In continents, crustal radioactivity accounts for a large fraction of the surface heat flux, such that the basal heat flux out of the convecting mantle is very small. Continental radioactivity plays no dynamic role and we may thus, as a first approximation, equate the heat loss of Earth to that of the oceans. For the sake of simplicity, we specify the heat flux using a half-space cooling model. Predictions from that model differ from observations at old ages only, and the difference represents a small fraction of the total: changing heat loss through seafloor older than 80 My by 30% (an overestimate) impacts the total planetary value by only 6%. Thus, we assume that $q=\lambda T_M/\sqrt{\pi\kappa\tau}$. For a rectangular plate of velocity U and width l, heat loss Q_P is

$$Q_P=\int_A q\mathrm{d}A=\int_0^{\tau_m} qlU\mathrm{d}\tau=2\frac{\lambda}{\sqrt{\pi\kappa}}T_M Ul\sqrt{\tau_m} \qquad [64]$$

where τ_m is the maximum plate age. The plate length is $L=U\tau_m$ and the rate of heat loss over the total area of the plate, $A_P=Ll$, is

$$Q_P=2A_P\frac{\lambda T_M}{\sqrt{\pi\kappa\tau_m}} \qquad [65]$$

This result holds for a plate where subduction occurs at the same age everywhere, such that the distribution of seafloor ages is a boxcar function (which shall be referred to as a 'rectangular' age distribution). Note that it was derived directly from eqn [64] independently of dynamic constraints on the spreading velocity. All the dynamic information needed is in the subduction age τ_m, which has not been determined yet.

It may be shown that all the other key variables of convection, plate velocity U and length L, depend on τ_m. To obtain a closure equation for this variable, a widely used argument is that the thermal boundary layer becomes unstable when a local Rayleigh number exceeds a threshold value. This leads to the 'parameterized' model with $\beta=1/3$, which is not satisfactory. A key piece of physics is missing, which would specify how τ_m changes when the mantle cools down. When the Earth was hotter, lithospheric plates were stiffer than today (Korenaga, 2006), but consequences for the subduction age τ_m or the plate length L are not obvious. One might imagine, for example, that the higher intrinsic stiffness of the lithosphere gets compensated by subducting younger and hence thinner plates, such that the total dissipation due to bending does not change much.

To specify the controls on the Earth's heat loss, one runs into two different types of difficulties. One must understand what determines the maximum plate age τ_m and one must go from a single convection cell to several cells with completely different characteristics. Even if we restrict ourselves to the major plates on Earth today, we cannot overlook the fact that both their size and spreading rate vary by one order of magnitude. In such conditions, the best strategy is to derive useful relationships without reference to dynamic arguments.

We use again the half-space cooling model that is sufficiently accurate for our purposes. The distribution of seafloor area can be expressed as a function of dimensionless age τ/τ_m:

$$\frac{\mathrm{d}A}{\mathrm{d}\tau}=C_A f\left(\frac{\tau}{\tau_m}\right) \qquad [66]$$

where C_A is the plate accretion rate. By construction, dimensionless function f goes from 1 for $\tau/\tau_m=0$ to 0 for $\tau/\tau_m=1$. Using the heat flux expression [30], the total oceanic heat loss is

$$Q_{oc}=A_o\frac{\lambda T_M}{\sqrt{\pi\kappa\tau_m}}\frac{\int_0^1\frac{f(u)}{\sqrt{u}}\mathrm{d}u}{\int_0^1 f(u)\mathrm{d}u}=A_o\frac{\lambda T_M}{\sqrt{\pi\kappa\tau_m}}\xi(f) \qquad [67]$$

where A_o is the total ocean surface and $\xi(f)$ a coefficient that depends on the dimensionless age distribution. For the present-day triangular age distribution,

$$Q_{oc}=\frac{8A_o}{3}\frac{\lambda T_M}{\sqrt{\pi\kappa\tau_m}} \qquad [68]$$

This has the same form as the previous result for a rectangular plate (eqn [65]), with a different proportionality constant.

Even though we have not used any dynamic argument to derive the heat loss equation, we can relate changes of heat loss to changes of spreading rate. The total oceanic area is

$$A_o=C_A\tau_m\int_0^1 f(u)\mathrm{d}u \qquad [69]$$

If we assume that the oceanic area and the age distribution remain constant, changes of C_A, the rate of plate generation, imply changes of τ_m, which in turn imply changes of heat loss at constant mantle temperature.

The oceanic heat loss can be equated with little error to the convective heat loss on Earth. Thus, eqns [65], [67], and [69] provide the most compact description of the physical controls on cooling. They avoid the problem of defining an 'average' plate with average velocity and length. They say nothing, however, about the maximum plate age τ_m.

7.06.7.5 Vagaries of Seafloor Spreading and Heat Loss

We have taken great care in calculating the various items in the energy budget but have been dealing with a record of oceanic spreading and geologic activity that is very short compared to the planet's age. Thus, we must evaluate whether our estimates are truly representative of the Earth's secular evolution or not.

In eqn [67], one coefficient that can vary significantly over short timescales is $\xi(f)$, which depends on the dimensionless

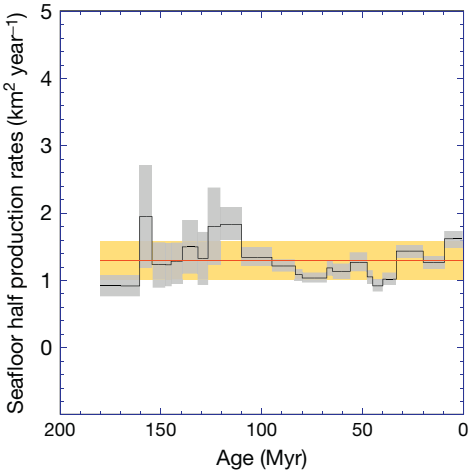


Figure 18 Variation of seafloor spreading rates over the last 180 My accounting for oceanic ridges that got subducted in the Pacific and Tethys Oceans. The black line is the half spreading rate and the gray area represents the uncertainty. The red line is the average value. Adapted from Cogné J-P and Humler E (2004) Temporal variation of oceanic spreading and crustal production rates during the last 180 My. *Earth and Planetary Science Letters* 227: 427–439.

distribution of seafloor ages. A few authors have attempted to calculate spreading rates in the past. The last 180 My has seen the closure of the Tethys Ocean and the subduction of several ridges in the paleo-Pacific Ocean (Engebretson et al., 1984). Accounting for those, the total seafloor generation rate has not changed significantly for the past 180 My according to Cogné and Humler (2004) (**Figure 18**). Fluctuations in seafloor spreading may occur on a larger timescale. Subduction of young seafloor occurs mostly at the edge of continents and may be due to the complex shapes of ocean–continent boundaries. With all continents assembled in a single landmass, the large continuous oceanic domain imposes less constraints on spreading and subduction. In other words, the present-day distribution of subduction zones may be a transient feature associated with the breakup of Gondwanaland (Seton et al., 2009). Coltice et al. (2012) had recently reproduced the triangular oceanic age distribution with numerical calculations of convection in a spherical Earth with platelike behavior and rigid blocks acting as continents. They found that, when continents are assembled, ocean–ocean subduction dominates and generates the same rectangular age distribution as classical convection systems where subduction is entirely controlled by buoyancy.

The assembly and breakup of supercontinents may affect the distribution of seafloor ages in the oceans and hence the Earth's heat loss. Such global reorganization occurs over some characteristic time τ_W. Allègre and Jaupart (1985) had related this time to the 'mean free path' of continents, such that continents sweep the whole surface of the Earth and necessarily run into one another. They obtained $\tau_W \approx 400$ My for present-day spreading rates and distribution of continents, which is less than the thermal adjustment time τ_p (eqn [62]). τ_W varies as a function of continental area and drift velocity and has probably been larger in the past when continents occupied a smaller fraction of the Earth's surface. Geologic data support such an increase of τ_W (Hoffman, 1997). Note that this observation runs against the intuitive notion that plates moved faster in the past. If the rate of heat loss of the Earth depends on the distribution of continents, it oscillates on a timescale τ_W over a long-term decreasing trend. We now attempt to estimate how large these oscillations can be.

Convective systems that are not constrained by lateral boundaries lead to age distributions that are almost rectangular (see **Appendix H**). In this case, subduction occurs at the same age everywhere and parameter $\xi(f)$ is equal to 2, as shown by eqn [65], which is 25% less than the value for the triangular distribution (8/3). This decreases the estimated oceanic heat loss by about 8 TW, which amounts to about half of the difference between present heat loss and heat production in the mantle (**Table 12**). Assuming for the sake of argument that the lifetime of a supercontinent is as long as the Wilson cycle and that these cycles are accompanied by changes of the age distribution of seafloor, the time-averaged oceanic heat loss could be 12.5% (i.e., 4 TW) less than the estimate of **Table 12**. A similar result was obtained by Grigné et al. (2005) using numerical models with many interacting convection cells of variable wavelength. As shown by Labrosse and Jaupart (2007), such fluctuations have a small influence on the Earth's thermal evolution because they are too small to cause significant changes of the planet's bulk internal energy and temperature. They do, however, impact arguments on the Urey ratio and the global energy budget.

Another variable that enters the heat loss expression (eqn [67]) is the total area of oceans. So far, we have assumed that it remains constant through time, but changes of oceanic area that are not related to rigid plate tectonics are likely to occur due to zones of diffuse deformation. Such zones are found in both oceans and continents and presently account for $\approx 15\%$ of the Earth's surface (Gordon, 2000). In continents, extension occurs at the expense of oceans, whereas shortening increases the oceanic area. These zones are usually very active and characterized by high heat flux, as in the Basin and Range Province, for example. Assuming that the average heat flux over such zones is equal to that in the Basin and Range Province (105 mW m^{-2}), the net effect on the global heat loss is small because this heat flux is $\approx$ the average oceanic one.

Yet another type of transient may be due to alternating phases of enhanced and subdued hot spot activity. The magnitude of the induced heat loss variations is likely to be small given the relatively modest contribution of mantle plumes to the total heat loss (<4 TW).

Three types of transient phenomena are likely to induce fluctuations of the Earth's rate of heat loss. Thus, one should allow for small, but significant, departures from today's energy budget in studies of the Earth's secular thermal evolution. These departures are not large enough to explain the difference between the present and the average values of the secular cooling rate, however, and we evaluate the role that may be played by the Earth's core.

7.06.7.6 Heat Flow from the Core

Collapsing the thermal evolution of the Earth into a single energy equation (eqn [53]) explicitly assumes that only one temperature is sufficient to describe internal energy changes or

equivalently that the core and the mantle evolve at proportional rates. There are reasons to think otherwise: as discussed in Section 7.06.5.3, the core has probably been cooling by more than 750 K, and possibly by as much as 1000 K, in 4 Gy, whereas the mantle has cooled by <200 K over the same interval. Let us consider how this affects the thermal evolution of the Earth.

Allowing for imperfect coupling between mantle and core, the global energy balance must allow for two different temperature evolutions:

$$M_m\langle C_m\rangle\frac{dT_m}{dt}+M_c\langle C_c\rangle\frac{dT_c}{dt}=-Q+H \qquad [70]$$

where, for simplicity, the effective heat capacities for the core and mantle, $\langle C_c\rangle$ and $\langle C_m\rangle$, respectively, account for isentropic temperature gradients and the consequences of inner core growth. M_m and M_c are the mantle and core masses, respectively. Taking average cooling rates of $50\,K\,Gy^{-1}$ and $250\,K\,Gy^{-1}$ for mantle and core, respectively, and $H=20$ TW, we find that the present-day heat loss is $Q=43$ TW, close to the observations. With the additional degree of freedom brought by core cooling, one can readily reconcile the present-day energy budget with available constraints on the mantle cooling rate and heat production. In other words, the difference between the present-day cooling rate of Earth ($98\,K\,Gy^{-1}$, Table 11) and the secular average value for the mantle ($50\,K\,Gy^{-1}$, Section 7.06.6) may be due in part to imperfect thermal coupling between mantle and core.

7.06.7.7 Summary

This section has emphasized that it is not easy to get around the powerful constraint of the energy budget. Most theoretical models for the cooling of Earth fail to account for the present-day imbalance between heat loss and heat production. This may be due in part to a particular convection phase that temporarily enhances heat loss compared to the long-term evolution. Past mantle temperature changes are documented in a variety of ways and provide key evidence on the Earth's convective engine. In particular, they imply that, in the past, the bulk rate of heat loss has changed less rapidly than that of heat production. The assumption that temperatures in the core and the mantle evolve at proportional rates is a gross simplification, which is difficult to reconcile with recent results on core thermal properties and geodynamo requirements.

7.06.8 Conclusions

Studies of the thermal evolution of the Earth are as old as physics and remain central to geology, for they deal with the energy that drives all geologic processes. The heat budget of the mantle can be established with a reasonable accuracy ($\approx 20\%$), thanks to the tremendous improvements in our knowledge of physical properties and data coverage. The Urey number is the ratio of heat production to heat loss, two imperfectly known quantities whose estimates are summarized in Table 12. Our current estimates for the Urey number are in the range 0.12–0.49, which rules out most available cooling models. Heat loss does not follow the decay of radiogenic heat production, and the time lag is on the order of the age of the Earth (see Section 7.06.7.1).

The present mantle energy budget implies a secular cooling rate in the range of 7–210 K Gy^{-1}. Over a long timescale, the average value for the cooling rate estimated from geologic constraints appears to be at the low end of this range (50 K Gy^{-1}). There is no reason to assume that the cooling rate has remained constant through time. Both geologic data and physical constraints on the thermal structure of the early Earth indicate that the cooling rate has increased as the planet got older (see Section 7.06.7.5). Plate tectonics is a regime of mantle convection attainable only in a state of subsolidus rheology that is with at most 40% melt. Thus, independent constraints on the cooling rate come from considerations of early Earth conditions. The most recent phase diagrams for the mantle indicate that the 40% melt threshold is reached when the potential temperature is about 200 K higher than at present (Litasov and Ohtani, 2002; Zerr et al., 1998; see also Boehler, vol. 2). If plate tectonics has been operating since the end of the magma ocean, the total amount of cooling is known and the average cooling rate cannot be more than about 50 K Gy^{-1}.

Appendix A Contraction of the Earth due to Secular Cooling

The planet contracts as it cools down. This induces changes of gravity that themselves induce changes of pressure and density. Here, we derive an approximate solution for a homogeneous planet in order to show how this affects the contraction rate and to identify the control variables.

Assuming spherical symmetry, governing equations are as follows:

$$\frac{1}{r^2}\frac{d}{dr}(r^2g)=4\pi G\rho \qquad [A.1]$$

$$\frac{dP}{dr}=-\rho g \qquad [A.2]$$

where G is the gravitational constant and P the pressure. From the equation of state, we deduce that

$$\frac{d\rho}{dr}=\frac{d\rho}{dP}\frac{dP}{dr}=\frac{\rho}{K_s}\frac{dP}{dr} \qquad [A.3]$$

$$=-\frac{\rho^2 g}{K_s} \qquad [A.4]$$

where K_s is the isentropic bulk modulus. For simplicity, the bulk modulus is assumed to be constant. Using radius R as length scale and the density at the top (defined more precisely later) ρ_T as density scale, the proper gravity scale is

$$[g]=G\rho_T R \qquad [A.5]$$

Equations can now be written using dimensionless variables:

$$\frac{1}{r^2}\frac{d}{dr}(r^2g)=4\pi\rho \qquad [A.6]$$

$$\frac{d\rho}{dr}=-\epsilon\rho^2 g \qquad [A.7]$$

where ε is a dimensionless number that provides a measure of the magnitude of density changes due to pressure:

$$\varepsilon = \frac{\rho_T[g]R}{K_s} = \frac{G\rho_T^2 R^2}{K_s} \quad [A.8]$$

We consider that the radial temperature profile is isentropic and neglect the upper thermal boundary layer that does not contribute much to the total mass and energy. Thus, reference density ρ_T is the mantle density at the atmospheric pressure, that is, at the potential temperature of the isentropic radial profile. For what follows, the key point is that ρ_T is not affected by internal pressure changes due to contraction.

For $\rho_T = 3.3\times10^3\,\mathrm{kg\,m^{-3}}$, $R = 6370$ km, and a value of 150 GPa for K_s, $\varepsilon \approx 0.2$. This is small and we expand all the variables in series of ε. We find

$$\rho = 1 + \varepsilon\left[-\frac{4\pi}{6}\left(r^2-1\right)\right] + \cdots \quad [A.9]$$

$$g = \frac{4\pi}{3}r + \varepsilon\left[\frac{16\pi^2}{90}r\left(5-3r^2\right)\right] + \cdots \quad [A.10]$$

To first order in ε, the mass of the planet is

$$M = \frac{4}{3}\pi\rho_T R^3\left(1+\frac{4\pi}{15}\varepsilon\right) = \frac{4}{3}\pi\rho_T R^3\left(1+\frac{4\pi}{15}\frac{G\rho_T^2R^2}{K_s}\right) \quad [A.11]$$

Writing that mass is conserved when ρ_T and R change, we obtain

$$\frac{\Delta R}{R} = -\frac{1}{3}\frac{\Delta\rho_T}{\rho_T}\left(1+\frac{16\pi}{45}\frac{G\rho_T^2R^2}{K_s}\right) \quad [A.12]$$

By definition, surface density ρ_T is calculated at the atmospheric pressure and hence depends on temperature only, such that

$$\frac{\Delta\rho_T}{\rho_T} = -\alpha\Delta T \quad [A.13]$$

where α is the thermal expansion coefficient at the atmospheric pressure. The end result is therefore

$$\frac{\Delta R}{R} = \frac{\alpha\Delta T}{3}\left(1+\frac{16\pi}{45}\frac{G\rho_T^2R^2}{K_s}\right) \quad [A.14]$$

This shows that contraction is enhanced by the change in gravity field.

Appendix B Gravitational Energy Changes

Here, we present a demonstration for changes of the bulk gravitational energy that was made by Paul Roberts (pers. comm.). For a sphere of uniform density and radius R,

$$E_g = -\frac{16\pi^2}{15}G\rho^2R^5 \quad [B.1]$$

For a contracting object with moving material boundaries, it is convenient to employ the Lagrangian variables. Because of mass conservation, the most convenient variable is the mass in a sphere of radius r:

$$M(r) = 4\pi\int_0^r \rho(r)r^2\mathrm{d}r = \int_0^r \rho(r)\mathrm{d}V \quad [B.2]$$

Thus,

$$g(r) = \frac{GM(r)}{r^2} \quad [B.3]$$

The gravitational energy in a material volume bounded by spheres of radii r_1 and r_2 is

$$E_{12} = -\int_{r_1}^{r_2}\rho g r\mathrm{d}V = -\int_{M_1}^{M_2} rg\mathrm{d}M = -\int_{M_1}^{M_2}\frac{GM}{r}\mathrm{d}M \quad [B.4]$$

The Lagrangian derivative is easily calculated using the variable M:

$$\frac{\mathrm{d}E_{12}}{\mathrm{d}t} = \int_{M_1}^{M_2}\frac{\mathrm{d}r}{\mathrm{d}t}\frac{GM}{r^2}\mathrm{d}M \quad [B.5]$$

For this calculation, the velocity field is limited to contraction, such that $\mathrm{d}r/\mathrm{d}t = v_c$. Thus,

$$\frac{\mathrm{d}E_{12}}{\mathrm{d}t} = \int_{M_1}^{M_2} v_c\frac{GM}{r^2}\mathrm{d}M = -\int_{r_1}^{r_2}\rho(\mathbf{v}_c\cdot\mathbf{g})\mathrm{d}V \quad [B.6]$$

which is the integral form of eqn [17] in the main text.

Appendix C Viscous Dissipation

The developments that follow stem in part from comments by Yanick Ricard. Viscous dissipation involves deviatoric stresses and departures from hydrostatic equilibrium. For steady-state conditions, Alboussière and Ricard (2013) had thoroughly examined the various assumptions and simplifications that have been made over the years and had come up with several alternative forms of the dissipation equation. In transient thermal conditions, the problem becomes more complicated as one must deal with changes of the hydrostatic base state. The base state is defined by the azimuthal averages of temperature and pressure and follows an isentrope, defined by its entropy $\bar{s}$, temperature $\overline{T}$, pressure $\overline{p}$, and density $\overline{\rho}$. By definition of the isentrope, $\bar{s}$ does not vary with radial distance, such that

$$\rho T\frac{\partial\bar{s}}{\partial r} = \rho C_p\frac{\partial\overline{T}}{\partial r} - \alpha\overline{T}\frac{\partial\overline{p}}{\partial r} = 0 \quad [C.1]$$

From hydrostatics,

$$\frac{\partial\overline{p}}{\partial r} = -\overline{\rho}g \quad [C.2]$$

which leads to

$$\frac{\partial\overline{T}}{\partial r} = -\frac{\alpha g}{C_p}\overline{T} \quad [C.3]$$

which defines the isentropic radial temperature profile. $\bar{s}$ is a function of time only and is well suited to track changes of the base state as the planet is cooling down.

For convection in the solid Earth, inertial terms can be neglected in the momentum balance, with the consequence that changes of kinetic energy are also negligible in the mechanical energy balance equation [10]. Subtracting the hydrostatic balance equation [14] from the momentum balance, we get

$$0 = -\nabla(\delta p) - \nabla\cdot\boldsymbol{\sigma} + \delta\rho\mathbf{g} \quad [C.4]$$

where $\delta\rho$ and $\delta\rho$ are deviations from the azimuthal averages of pressure and density, respectively, and where we have neglected perturbations to the gravity field, which only appear in second-order terms. Introducing the temperature

perturbation, θ, the density perturbation is obtained by a Taylor expansion about the mean:

$$\frac{\delta\rho}{\overline{\rho}}=\frac{\delta p}{K_T}-\alpha\theta \qquad [C.5]$$

Introducing the isentropic bulk modulus K_S and Grüneisen parameter γ, such that $K_S=K_T(1+\gamma\alpha T)$, we obtain

$$0=-\nabla(\delta p)+\overline{\rho}\mathbf{g}\left[\frac{\delta p}{K_S}+\frac{\gamma\alpha T}{K_S}\delta p-\alpha\theta\right]-\nabla\cdot\boldsymbol{\sigma} \qquad [C.6]$$

which can be recast as

$$0=-\overline{\rho}\nabla\left(\frac{\delta p}{\overline{\rho}}\right)+\overline{\rho}\mathbf{g}\left[\frac{\gamma\alpha T}{K_S}\delta p-\alpha\theta\right]-\nabla\cdot\boldsymbol{\sigma} \qquad [C.7]$$

Using the entropy equation, we obtain

$$0=-\overline{\rho}\nabla\left(\frac{\delta p}{\overline{\rho}}\right)-\frac{\alpha\overline{\rho}T\mathbf{g}}{C_p}\delta s-\nabla\cdot\boldsymbol{\sigma} \qquad [C.8]$$

where δs is the entropy perturbation. Taking the scalar product with convective velocity $\mathbf{w}$, we deduce a modified form of the mechanical energy balance:

$$0=-\overline{\rho}\mathbf{w}\cdot\nabla\left(\frac{\delta p}{\overline{\rho}}\right)-\frac{\alpha\overline{\rho}T}{C_p}(\mathbf{w}\cdot\mathbf{g})\delta s-\mathbf{w}\cdot\nabla\cdot\boldsymbol{\sigma} \qquad [C.9]$$

Integrating this equation over the planet volume yields three separate integrals, which are evaluated separately. The first one on the right yields

$$\int_V\overline{\rho}\mathbf{w}\cdot\nabla\left(\frac{\delta p}{\overline{\rho}}\right)\mathrm{d}V=\int_A\delta p\,\mathbf{w}\cdot\mathrm{d}A-\int_V\frac{\delta p}{\overline{\rho}}\nabla\cdot(\overline{\rho}\mathbf{w})\mathrm{d}V \qquad [C.10]$$

The two contributions on the right-hand side are zero. For the first one, we use the fact that the pressure perturbation δp must vanish at the Earth's surface where pressure is held constant. For the second one, we rely on the continuity equation in the anelastic approximation, $\nabla\cdot(\overline{\rho}\mathbf{w})=0$ (Braginsky and Roberts, 1995). The third integral on the right of the mechanical energy balance equation [C.9] yields

$$\begin{aligned}\int_V\mathbf{w}\cdot\nabla\cdot\boldsymbol{\sigma}\mathrm{d}V&=\int_V(\mathbf{w}\cdot\nabla\cdot\boldsymbol{\sigma}-\nabla\cdot[\boldsymbol{\sigma}\cdot\mathbf{w}])\mathrm{d}V+\int_V\nabla\cdot[\boldsymbol{\sigma}\cdot\mathbf{w}]\mathrm{d}V\\&=\int_V\phi\mathrm{d}V+\int_A\mathbf{w}\cdot[\boldsymbol{\sigma}\cdot\mathrm{d}A]\end{aligned} \qquad [C.11]$$

The last term on the right of this equation is the work of shear stresses at the Earth's surface, which is negligible. Using these results, the mechanical energy equation may be reduced to

$$\int_V\overline{\phi}\mathrm{d}V=-\int_V\frac{\alpha\overline{\rho}T}{C_p}(\mathbf{w}\cdot\mathbf{g})\delta s\mathrm{d}V \qquad [C.12]$$

which specifies dissipation. Under a series of assumptions, it may be shown that the integral on the right can be reduced to (Alboussière and Ricard, 2013)

$$\int_V\frac{\alpha\overline{\rho}T}{C_p}(\mathbf{w}\cdot\mathbf{g})\delta s\mathrm{d}V=-\int_V\rho\alpha g\overline{w_r\theta}\mathrm{d}V \qquad [C.13]$$

where w_r is the radial velocity component. This integral can be interpreted as the work of buoyancy forces and has been called the 'buoyancy flux.' The power of the dissipation equation in this form is that it can be expressed as a function of the convective heat flux $\left(\rho C_p\overline{w_r\theta}\right)$ and can be related to heat loss directly. Alboussière and Ricard (2013) had emphasized its approximate nature, however, such that it cannot be applied to complex thermodynamic systems without care.

Dissipation enters the local energy balance (eqn [11]) but not the global energy budget for the planet as a whole (eqn [1]). To find out how this comes about, we separate entropy into the azimuthal average $\bar{s}$ and perturbation δs:

$$\overline{\rho}T\frac{\mathrm{D}s}{\mathrm{D}t}=\overline{\rho}T\frac{\mathrm{D}\bar{s}}{Dt}+\overline{\rho}T\frac{\partial\delta s}{\partial t}+\overline{\rho}T\mathbf{w}\cdot\nabla\delta s \qquad [C.14]$$

where we have relied on the anelastic approximation in using $\overline{\rho}$ instead of ρ (Braginsky and Roberts, 1995). The last term on the right can be rewritten as follows:

$$\overline{\rho}T\mathbf{w}\cdot\nabla\delta s=\nabla\cdot(\overline{\rho}T\delta s\mathbf{w})-\overline{\rho}(\mathbf{w}\cdot\nabla T)\delta s-T\delta s\nabla\cdot(\overline{\rho}\mathbf{w}) \qquad [C.15]$$

The third term on the right can be set to zero using mass conservation, as mentioned earlier. The first term on the right is written in conservative form and leads to zero upon integration over the volume. Thus,

$$\int_V\overline{\rho}T(\mathbf{w}\cdot\nabla\delta s)\mathrm{d}V=-\int_V\overline{\rho}(\mathbf{w}\cdot\nabla T)\delta s\mathrm{d}V \qquad [C.16]$$

Collecting terms in the energy balance and neglecting external energy transfers (in the ψ term), we find

$$\begin{aligned}\int_V\overline{\rho}T\frac{\mathrm{D}\bar{s}}{\mathrm{D}t}\mathrm{d}V+\int_V\overline{\rho}T\frac{\partial\delta s}{\partial t}\mathrm{d}V&=\int_V\overline{\rho}\mathbf{w}\cdot\left(\nabla T-\frac{\alpha T}{C_p}\mathbf{g}\right)\delta s\mathrm{d}V\\&\quad-\int_A\mathbf{q}\cdot\mathrm{d}A+\int_VH\mathrm{d}V\end{aligned} \qquad [C.17]$$

If we neglect second-order terms, the first term on the right involves temperature variations along the isentrope and cancels because of our assumption of an isentropic base state defined by eqn [C.3]. We now evaluate the second term on the left. Integrating over the volume involves surface integrals:

$$\int_V\overline{\rho}T\frac{\partial\delta s}{\partial t}\mathrm{d}V=\int_0^R\left(\int_A\overline{\rho}T\frac{\partial\delta s}{\partial t}\mathrm{d}A\right)\mathrm{d}r \qquad [C.18]$$

We split temperature into the azimuthal average $\overline{T}$ and perturbation θ and consider the two corresponding contributions. For the former, by definition of the azimuthal average, the surface integral over a spherical shell of radius r can be expressed as follows:

$$\int_A\overline{\rho}\overline{T}\frac{\partial\delta\bar{s}}{\partial t}\mathrm{d}A=\left(\overline{\rho}\overline{T}\frac{\partial\delta\bar{s}}{\partial t}\right)4\pi r^2 \qquad [C.19]$$

which is zero by definition of the entropy perturbation δs. The remaining term involving the temperature perturbation θ and the entropy perturbation δs is a second-order term that can be neglected as before.

The first term on the left of eqn [C.17] involves the entropy of the base state:

$$\overline{\rho}T\frac{\overline{\mathrm{D}s}}{\mathrm{D}t}=\overline{\rho}T\left(\frac{\partial\bar{s}}{\partial t}+\mathbf{w}\cdot\nabla\bar{s}\right)=\overline{\rho}T\left(\frac{\partial\bar{s}}{\partial t}\right) \qquad [C.20]$$

where we have used $\nabla\bar{s}=0$ by definition of the isentrope. Writing entropy variations with state variables T and ρ, we get

$$\overline{\rho}T\frac{\partial \overline{s}}{\partial t}=\overline{\rho}C_v\frac{\partial \overline{T}}{\partial t}-\frac{\alpha TK_T}{\rho}\frac{\partial \rho}{\partial t} \quad [C.21]$$

We therefore end up with an equation for temperature:

$$\int_V \overline{\rho}C_v\frac{\partial \overline{T}}{\partial t}dV=\int_V \frac{\alpha TK_T}{\rho}\frac{\partial \rho}{\partial t}dV-\int_A \mathbf{q}\cdot dA+\int_V H dV \quad [C.22]$$

where the first term on the right is the only remaining contribution of contraction and has been referred to as 'adiabatic heating.' This term is small for solid-state planets because of their small values of the thermal expansion coefficient ($\alpha T \ll 1$). If one uses an alternative form of the entropy equation with pressure and temperature as the two independent state variables, one deals with C_p instead of C_v. The former varies with temperature but the latter does not (at least in the Dulong–Petit–Debye limit). Using the thermodynamic identity $C_p=C_v(1+\gamma\alpha T)$, one can see that it is not self-consistent to deal with adiabatic heating and ignore variations of C_p at the same time (Alboussière and Ricard, 2013). From a purely practical point of view, however, as long as one does not introduce such simplifications into formal thermodynamic arguments, there is no harm in keeping negligible terms in an equation.

Appendix D Half-Space Cooling Model with Temperature-Dependent Properties

The general form of the one-dimensional heat equation is

$$\rho C_p\frac{\partial T}{\partial \tau}=\frac{\partial}{\partial z}\left(\lambda\frac{\partial T}{\partial z}\right) \quad [D.1]$$

where all properties depend on temperature T. In this section, the temperature for the half-space is initially T_M and temperature is set to zero at the surface: $T(z=0,\tau)=0$.

For constant physical properties, the temperature for the half-space is given by (Carslaw and Jaeger, 1959)

$$T(z,t)=T_M\mathrm{erf}\left(\frac{z}{2\sqrt{\kappa\tau}}\right) \quad [D.2]$$

and the surface heat flux is

$$q(0,t)=\frac{\lambda T_M}{\sqrt{\pi\kappa\tau}} \quad [D.3]$$

For temperature-dependent properties, Carslaw and Jaeger (1959) introduced the following variable:

$$\theta=\frac{1}{\lambda_0}\int_0^T \lambda dT \quad [D.4]$$

where λ_0 is thermal conductivity at the reference temperature $T=0$; the heat equation becomes:

$$\frac{1}{\kappa}\frac{\partial \theta}{\partial \tau}=\frac{\partial^2\theta}{\partial z^2} \quad [D.5]$$

where κ is the temperature-dependent thermal diffusivity. For the boundary and initial conditions of interest here, this equation can be written in terms of a similarity variable η:

$$\theta(z,t)=\Theta(\eta), \text{ with } \eta=z\tau^{-1/2} \quad [D.6]$$

$$-\frac{\eta}{2\kappa}\Theta'=\Theta'' \quad [D.7]$$

with the following boundary conditions:

$$\Theta(0)=0, \Theta(\infty)=\Theta_M, \Theta'(\infty)=\Theta''(\infty)=0 \quad [D.8]$$

The surface heat flux is thus

$$q=\lambda\frac{\partial T}{\partial z}_{z=0}=k_0\tau^{-1/2}\Theta'(0) \quad [D.9]$$

which is of the form $C_Q\tau^{-1/2}$ regardless of the specific functional form of ρ, C_p, and λ. This beautiful result was pointed out first by Lister (1977).

Appendix E Plate Models for the Oceanic Lithosphere

For a plate of thickness a initially at temperature T_M, with $T(z=0,\tau)=0$ and $T(z=a,\tau)=T_M$, the temperature for $0<z<a$ is given by

$$T(z,\tau)=T_M\left(\frac{z}{a}+\sum_{n=1}^{\infty}\frac{2}{n\pi}\sin\left(\frac{n\pi z}{a}\right)\exp\left(\frac{-n^2\pi^2\kappa\tau}{a^2}\right)\right) \quad [E.1]$$

and the surface heat flux is given by

$$q(0,t)=\frac{\lambda T_M}{a}\left(1+2\sum_{n=1}^{\infty}\exp\left(\frac{-n^2\pi^2\kappa\tau}{a^2}\right)\right) \quad [E.2]$$

which diverges at $\tau=0$. For $\tau \ll a^2/\kappa$, the heat flux

$$q(0,\tau)\approx\frac{\lambda T_M}{\sqrt{\pi\kappa\tau}} \quad [E.3]$$

With fixed heat flux at the base, $Q(a,\tau)=\lambda T_M/a$, the temperature of the plate is given by

$$T(z,\tau)=\frac{T_M z}{a}-\frac{4T_M}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)}\cos\left(\frac{(2n+1)\pi(z-a)}{2a}\right)\times\exp\left(\frac{-(2n+1)^2\pi^2\kappa\tau}{4a^2}\right)+\frac{8T_M}{\pi^2}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}\sin\left(\frac{(2n+1)\pi z}{2a}\right)\exp\left(\frac{-(2n+1)^2\pi^2\kappa\tau}{4a^2}\right) \quad [E.4]$$

and the surface heat flux is given by

$$\frac{q(0,t)}{\frac{\lambda T_M}{a}}=1+2\sum_{n=0}^{\infty}\exp\left(\frac{-(2n+1)^2\pi^2\kappa\tau}{4a^2}\right)+\frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-)^n}{(2n+1)}\exp\left(\frac{-(2n+1)^2\pi^2\kappa\tau}{4a^2}\right) \quad [E.5]$$

The surface heat flux for these three different boundary conditions is compared in Figure 19, with the same length scale $L=a$ for the half-space and the constant temperature boundary condition and $L=2a$ for the heat flux boundary condition.

Appendix F Differences Between Estimates of the Energy Budget

Table 14 compares our construction of the mantle energy budget with those proposed by Stacey and Davis (2008) and

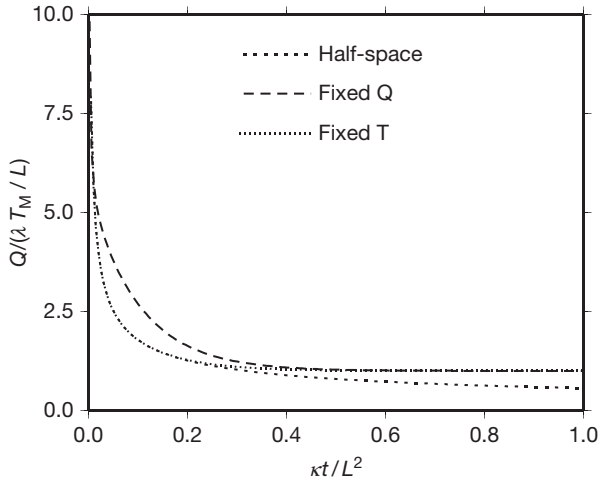


Figure 19 Normalized heat flux as a function of reduced time for different cooling models for the oceanic lithosphere. Length scale L is equal to a, the plate thickness, in the fixed temperature model and to 2a in the fixed flux model. In the half-space model, L cancels out. See **Appendix E**.

Table 14 Various estimates of the global budget

	Stacey and Davis (2008)	*Davies (1999)*	*This study*
Total heat loss	44	41	46
Continental heat production	8	5	8
Upper mantle		1.3	
Lower mantle		11–27	
Mantle heat production	20	12–28[a]	11
Latent heat–core differentiation	1.2	<1	
Mantle differentiation	0.6	0.3	0.2
Gravitational (thermal contraction)	3.1		
Tidal dissipation		0.1	0.1
Core heat loss	3.5	5	11
Mantle cooling	8	9[b]	16
Present Urey ratio	0.64	0.3–0.68	0.12–0.49

[a]Lower mantle heat production is variable and calculated to fit the mantle cooling rate.
[b]Mantle cooling is fixed.

Davies (1999). Although the total energy budget is almost identical, there are major differences in the breakdown of the budget between different items. These differences originate in the various assumptions made as well as in the objectives that are sought. Our estimate uses a slightly different total oceanic area that reflects our better knowledge of continental margins. It also includes the contribution of hot spots, which is not accounted for by bulk lithosphere cooling. Finally, it allows for a small amount of radioelements in the subcontinental lithospheric mantle and takes advantage of our improved constraints on crustal heat production.

For Stacey and Davis (2008), the total heat production (28 TW) is significantly higher than the value of 20 TW for BSE and almost equal to that of the nondepleted chondritic model of Birch (1965). It seems that these authors have added the crustal heat production to BSE. The core heat loss is low because it is assumed identical to the heat carried by hot spots. Stacey and Davis (2008) also assumed that all the gravitational energy released by thermal contraction (3.1 TW) goes to heat.

In Davies (1999), the secular cooling of the mantle is assumed to be fixed by petrologic data and the lower mantle heat production is the variable that is adjusted to balance the budget when all the other variables are fixed. Core cooling is also assumed identical to the total heat flux from hot spots. Upper mantle heat production is known to be low (from samples of the mantle carried to the surface). Lower mantle is assumed to be a mixture between a depleted chondritic composition, which would give 11 TW, and a MORB (!?) like component, which would yield 27 TW. The ratio of those two components is adjusted to balance the budget. (Note that depleted mantle should give only 6 TW and the MORB-like component seems to have the same heat production as that of chondrites.)

The global energy budget of the Earth was one of the arguments used by Kellogg et al. (1999) to propose that the lowermost mantle forms a distinct reservoir with chondritic concentration in radioelements, the 'abyssal layer.' The heat production in the depleted MORB mantle (i.e., the source of depleted MORBs) is $\approx 0.6\,\mathrm{pW\,kg^{-1}}$. Assuming that this is representative of the whole mantle, the total mantle heat production would only be 2.5 TW, that is, much less than the ≈13 TW obtained by removing the crustal heat production from BSE. (Note that Kellogg et al. (1999) used 31 TW for the total heat production in a 'chondritic' Earth.)

Appendix G Average Thermal Structure and Temperature Changes in Upwellings and Downwellings

The reference vertical temperature profile is often called 'adiabatic,' which is misleading. Here, we recapitulate the definitions and introduce two different reference temperature profiles. The equation for the entropy per unit mass, s, is

$$\rho T \frac{\mathrm{D}s}{\mathrm{D}t} = \rho C_\mathrm{p} \frac{\mathrm{D}T}{\mathrm{D}t} - \alpha T \frac{\mathrm{D}p}{\mathrm{D}t} \qquad [\mathrm{G}.1]$$

$$= -\nabla \cdot \mathbf{q} + \phi + H \qquad [\mathrm{G}.2]$$

Note that this shows that entropy is not conserved due to irreversible dissipation and radioactive decay. Density changes due to temperature have a small impact on pressure and dynamic pressure variations are small compared to the hydrostatic pressure. Thus,

$$\frac{\mathrm{D}p}{\mathrm{D}t} \approx -\rho g w \qquad [\mathrm{G}.3]$$

where w is the radial velocity component.

From these equations, we may deduce the isentropic temperature profile, such that $\mathrm{D}s/\mathrm{D}t = 0$. In the interior of the convecting system, far from the upper and lower boundaries, the dominant velocity component is the vertical. Assuming steady state and using eqn [G.1],

$$\rho C_\mathrm{p} w \frac{\mathrm{d}T_\mathrm{S}}{\mathrm{d}r} = -\alpha T_\mathrm{S} \rho g w \qquad [\mathrm{G}.4]$$

where T_S stands for the isentropic temperature profile. Thus,

$$\frac{dT_S}{dr} = -\frac{\alpha g}{C_p} T_S \qquad [G.5]$$

This is close to the vertical profile of the azimuthally averaged temperature in a steady-state well-mixed convective system with no internal heat production and negligible viscous dissipation.

The isentropic temperature gradient derived earlier provides a convenient *reference* profile that illustrates the role of compressibility. However, it is a poor approximation for the temperature path followed by a rising (or sinking) mantle parcel. We may consider for simplicity that such a parcel does not exchange heat with its surroundings, which is a good approximation for the broad return flow away from subduction zones. In this case, we set $\mathbf{q} = 0$, use the same approximation as before for pressure, and obtain

$$\rho C_p \frac{DT}{Dt} \approx \rho C_p \left(\frac{\partial T}{\partial t} + w \frac{\partial T}{\partial r} \right) = -(\alpha T) \rho g w + H + \phi \qquad [G.6]$$

The material parcel's temperature changes due to radiogenic heat production and dissipation and due to the work of pressure forces. This may be recast as follows:

$$\frac{\partial T}{\partial r} = -\frac{\alpha g}{C_p} T + \frac{1}{w} \left(\frac{\phi + H}{\rho C_p} - \frac{\partial T}{\partial t} \right) \qquad [G.7]$$

where one should note that secular cooling acts in the same direction as internal heat production. The radial temperature gradient from eqn [G.7] differs from the isentropic value by about 30%. Here, we see the importance of knowing precisely the secular cooling rate $\partial T/\partial t$.

Appendix H Seafloor Age Distribution as Seen from Models of Mantle Convection

We present here a few examples of temperature fields obtained in numerical models of convection and compute the seafloor age distribution they would provide for comparison with that

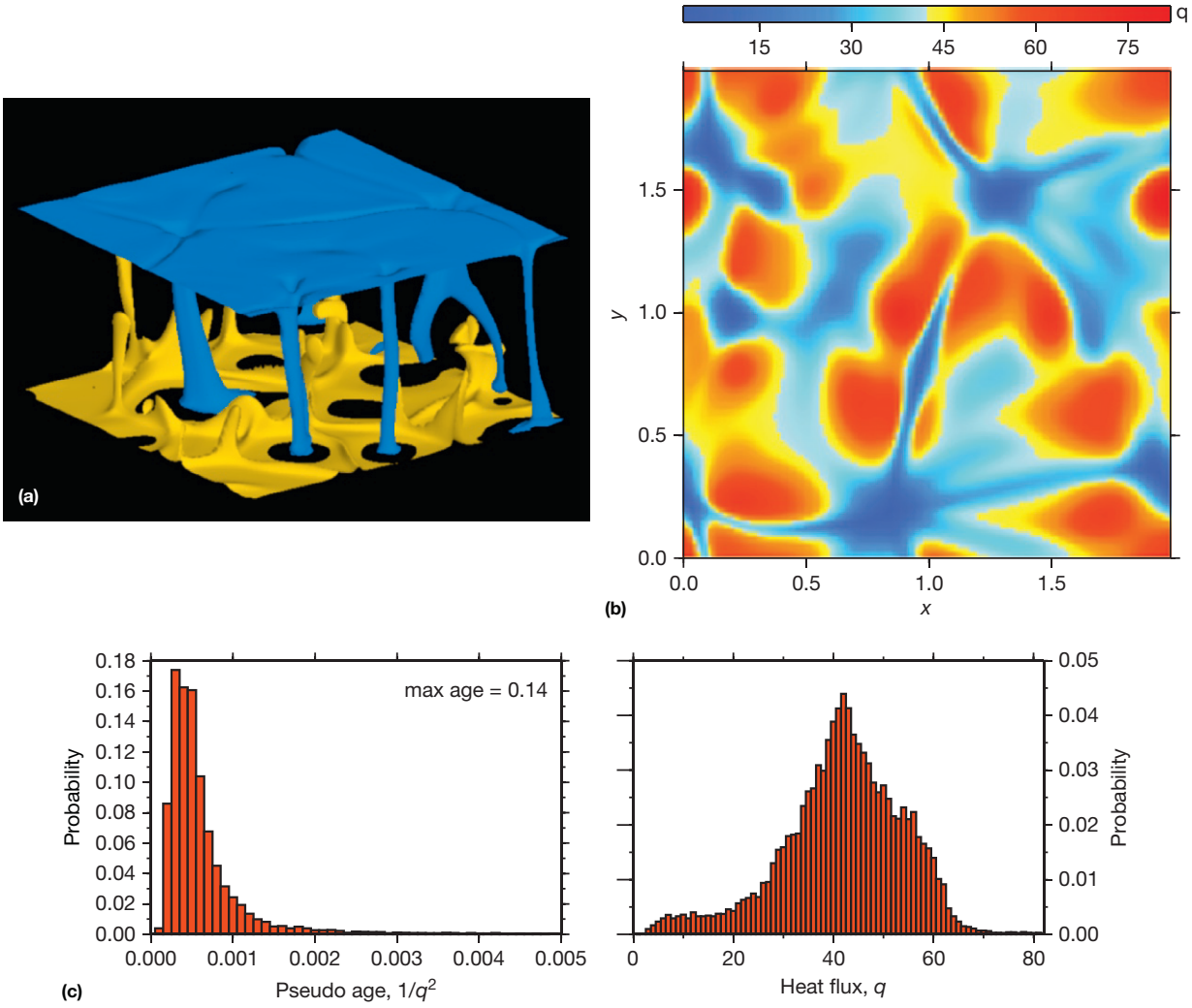


Figure 20 Snapshot of temperature (a), surface heat flux (b), and the corresponding pseudo-age distribution (c) in a convection system with a Rayleigh number $Ra = 10^7$ and an internal heating rate $H = 20$. Note that the age distribution is far from that of the Earth's oceans, which is triangular. From the model of Labrosse (2002). See **Appendix H** for details.

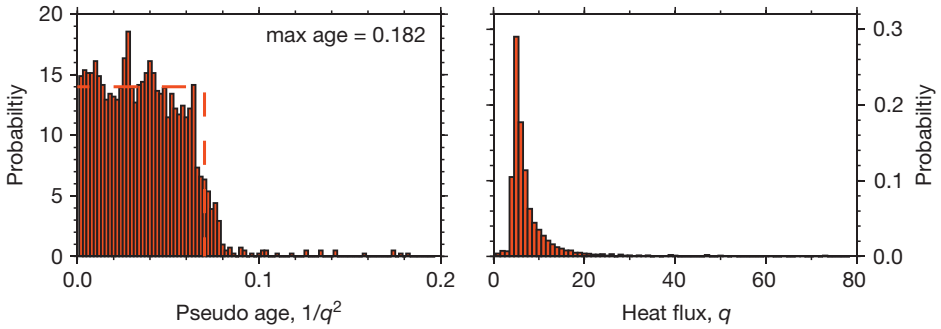


Figure 21 Distribution of surface heat flux (q, right) and pseudo age ($1/q^2$) for a model with platelike behavior, corresponding to **Figure 3**. The age distribution is close to a rectangular one (dashed line). See **Appendix H** for details. From the model of Grigne et al. (2005).

of Earth today. In all cases presented, convection is set by imposing a destabilizing temperature difference ΔT between top and bottom and a uniform internal heating rate per unit volume H. This defines two dimensionless parameters (e.g., Sotin and Labrosse, 1999), the Rayleigh number Ra and the dimensionless internal heating rate H^*:

$$Ra = \frac{g\alpha\Delta T D^3}{\kappa v_0}; \quad H^* = \frac{HD^2}{k\Delta T} \qquad [H.1]$$

where D is the thickness of the system and v_0 is the dynamic viscosity at some reference point. The other symbols are the same as in the rest of the chapter. A third dimensionless number is the Prandtl number $Pr = v_0/\kappa$, which is of order 10^{23} in the mantle and is accordingly set to infinity. The viscosity of mantle material is known to vary with temperature, pressure, mineral phase, etc. To reproduce the plate tectonics regime of mantle convection, a complex rheological law must be used (see **Chapters 7.01** and **7.03**). This introduces yet another set of parameters. For example, the pseudoplastic rheology used by Tackley (2000) is defined by an effective viscosity

$$\eta_{\text{eff}} = \min\left[\eta(T), \frac{\sigma_y}{2\dot{\varepsilon}}\right] \qquad [H.2]$$

where σ_y is a yield stress and $\dot{\varepsilon} = \sqrt{\dot{\varepsilon}_{i,j}\dot{\varepsilon}_{i,j}}/2$ the second invariant of the strain rate tensor. Viscosity is temperature-dependent:

$$\eta(T) = \exp\left[\frac{23.03}{T+1} - \frac{23.03}{2}\right] \qquad [H.3]$$

This specific rheological law was used to obtain the results of **Figure 3** (Grigné et al., 2005).

In numerical models of mantle convection, determining the age of each point at the surface requires extensive calculations with markers. To alleviate this difficulty, we use the fact that the surface heat flux is due to conductive cooling and define a pseudo age as follows:

$$q \propto \frac{1}{\sqrt{\tau}} \qquad [H.4]$$

We thus use heat flux values to determine the age distribution.

The first configuration tested is that of isoviscous Rayleigh–Bénard convection with internal heating. Using a snapshot from Labrosse (2002), we get a distribution of pseudo ages that is peaked at low values and exhibits a roughly exponential tail at large ages (**Figure 20**). This reflects that, with a large amount of internal heating, convective flow is dominated by strong cold plumes and passive return flow in most of the interior. There are weak hot plumes in the lower part of the domain, but they do not contribute much to the heat flux, in a manner reminiscent of that on Earth. The distribution of surface heat flux is due to the distributed return flow.

A somewhat more realistic model relying on the pseudoplastic rheology has been developed by Grigné et al. (2005) and a snapshot is shown in **Figure 3**. The distribution of surface heat flux and pseudo age for this model is given in **Figure 21**. The platelike behavior is such that the distribution of heat flux is peaked at low values and the pseudo-age distribution is approximately rectangular, reflecting the fact that all plates have similar sizes and velocity.

Acknowledgments

Denis Andrault, David Bercovici, Jean-Pascal Cogné, Earl Davis, Steve Dye, Mark Jellinek, Jun Korenaga, Charlie Langmuir, Paul Roberts, and Gerald Schubert provided us with much information and many ideas. A thorough review by Yanick Ricard clarified important thermodynamic aspects.

References

Abbott D, Burgess L, Longhi J, and Smith WHF (1994) An empirical thermal history of the Earth's upper mantle. *Journal of Geophysical Research* 99: 13835.

Abe Y (1993) Physical state of the very early Earth. *Lithos* 30: 223–235.

Abe Y (1997) Thermal and chemical evolution of the terrestrial magma ocean. *Physics of the Earth and Planetary Interiors* 100: 27–39.

Alboussière T and Ricard Y (2013) Reflections on dissipation associated with thermal convection. *Journal of Fluid Mechanics* 751: R1.

Alfé D, Gillan MJ, and Price GD (2002) Ab initio chemical potentials of solid and liquid solutions and the chemistry of the earth's core. *Journal of Chemical Physics* 116: 7127–7136.

Allègre CJ and Jaupart C (1985) Continental tectonics and continental kinetics. *Earth and Planetary Science Letters* 74: 171–186.

Andrault D, Bolfan-Casanova N, Nigro GL, Bouhifd MA, Garbarino G, and Mezouar M (2011) Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history. *Earth and Planetary Science Letters* 304: 251–259.

Andrault D, Muñoz M, Bolfan-Casanova N, et al. (2010) Experimental evidence for perovskite and post-perovskite coexistence throughout the whole D″ region. *Earth and Planetary Science Letters* 293: 90–96.
Andrault D, Petitgirard S, Nigro GL, et al. (2012) Solid–liquid iron partitioning in Earth's deep mantle. *Nature* 487: 354–357.
Anzellini S, Dewaele A, Mezouar M, Loubeyre P, and Morard G (2013) Melting of iron at Earth's inner core boundary based on fast X-ray diffraction. *Science* 340: 464–466.
Armitage JJ, Jaupart C, Fourel L, and Allen PA (2013) The instability of continental passive margins and its effect on continental topography and heat flow. *Journal of Geophysical Research* 118: 1817–1836.
Ashwal LD, Morgan P, Kelley SA, and Percival J (1987) Heat production in an Archean crustal profile and implications for heat flow and mobilization of heat producing elements. *Earth and Planetary Science Letters* 85: 439–450.
Barreyre T, Escartín J, Garcia R, Cannat M, Mittelstaedt E, and Prados R (2012) Structure, temporal evolution, and heat flux estimates from the Lucky Strike deep-sea hydrothermal field derived from seafloor image mosaics. *Geochemistry, Geophysics, Geosystems* 13: 4007.
Becker TW, Conrad CP, Buffett BA, and Müller RD (2009) Past and present seafloor age distributions and the temporal evolution of plate tectonic heat transport. *Earth and Planetary Science Letters* 278: 233–242.
Bender PL, et al. (1973) The Lunar Laser Ranging Experiment. *Science* 182: 229–238.
Birch F (1965) Speculations on the thermal history of the Earth. *Geological Society of America Bulletin* 76: 133–154.
Birch F, Roy RF, and Decker ER (1968) Heat flow and thermal history in New England and New York. In: An-Zen E (ed.) *Studies of Appalachian Geology*, pp. 437–451. New York: Wiley-Interscience.
Bird RB, Stewart WE, and Lightfoot EN (1960) *Transport Phenomena*. New York: Wiley.
Bodell JM and Chapman DS (1982) Heat flow in the north central Colorado plateau. *Journal of Geophysical Research* 87: 2869–2884.
Boillot G, Grimaud S, Mauffret A, et al. (1980) Ocean–continent boundary off the Iberian margin: A serpentinite diapir west of the Galicia Bank. *Earth and Planetary Science Letters* 48: 23–24.
Bonneville A, Von Herzen RP, and Lucazeau F (1997) Heat flow over Reunion hot spot track: Additional evidence for thermal rejuvenation of oceanic lithosphere. *Journal of Geophysical Research* 102: 22731–22748.
Borexino Collaboration Group (Bellini, G., 88 Collaborators) (2013) Measurement of geo-neutrinos from 1353 days of Borexino. *Physics Letters B* 722: 295–300.
Borexino Collaboration Group (Bellini, G., 89 Collaborators) (2010) Measurement of geo-neutrinos from 1353 days of Borexino. *Physics Letters B* 687: 299–304.
Boyet M and Carlson RW (2005) ^{142}Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.
Boyet M and Gannoun A (2013) Nucleosynthetic Nd isotope anomalies in primitive enstatite chondrites. *Geochimica et Cosmochimica Acta* 121: 652–666.
Bradley DC (2008) Passive margins through earth history. *Earth-Science Reviews* 91: 1–26.
Brady RJ, Ducea MN, Kidder SB, and Saleeby JB (2006) The distribution of radiogenic heat production as a function of depth in the Sierra Nevada Batholith, California. *Lithos* 86: 229–244.
Braginsky SI and Roberts PH (1995) Equations governing convection in Earth's core and the geodynamo. *Geophysical and Astrophysical Fluid Dynamics* 79: 1–97.
Breuer D and Spohn T (1993) Cooling of the earth, Urey ratios, and the problem of potassium in the core. *Geophysical Research Letters* 20: 1655–1658.
Calves G, Schwab AM, Huuse M, Clift PD, and Inam A (2010) Thermal regime of the northwest Indian rifted margin – Comparison with predictions. *Marine and Petroleum Geology* 27: 1133–1147.
Cameron AGW (2001) From interstellar gas to the Earth–Moon system. *Meteoritics and Planetary Science* 36: 9–22.
Canup RM (2004) Simulations of a late lunar-forming impact. *Icarus* 168: 433–456.
Canup RM and Righter K (eds.) (2000) *Origin of the Earth and Moon*. Tucson, AZ: University of Arizona Press.
Carlson RL and Johnson HP (1994) On modeling the thermal evolution of the oceanic upper mantle: An assessment of the cooling plate model. *Journal of Geophysical Research* 99: 3201–3214.
Caro G and Bourdon B (2010) Non-chondritic Sm/Nd ratio in the terrestrial planets: Consequences for the geochemical evolution of the mantle–crust system. *Geochimica et Cosmochimica Acta* 74: 3333–3349.
Caro G, Bourdon B, Birck J-L, and Moorbath S (2003) ^{146}Sm–^{142}Nd evidence from Isua metamorphosed sediments for early differentiation of the Earth's mantle. *Nature* 423: 428–432. http://dx.doi.org/10.1038/nature01668.
Caro G, Bourdon B, Halliday AN, and Quitte G (2008) Super-chondritic Sm/Nd ratios in mars, the earth and the moon. *Nature* 452: 336–339.
Carslaw HS and Jaeger JC (1959) *Conduction of Heat in Solids*, 2nd edn. Oxford: Clarendon Press.
Catalli K, Shim S-H, and Prakapenka VB (2009) Thickness and Clapeyron slope of the post-perovskite boundary. *Nature* 462: 782–785.
CCGM/CGMW (2000) *Geological Map of the World at 1:25 000 000*, 2nd edn. CD Rom, http://ccgm.free.fr/cartes_monde_gb.html.
Cermak V and Bodri L (1986) Two-dimensional temperature modelling along five East-European geotraverses. *Journal of Geodynamics* 5: 133–163.
Chapman DS and Pollack HN (1974) 'Cold spot' in West Africa: Anchoring the African plate. *Nature* 250: 477–478.
Christensen UR (1984) Heat transport by variable viscosity convection and implications for the Earth's thermal evolution. *Physics of the Earth and Planetary Interiors* 35: 264–282.
Christensen UR (1985) Thermal evolution models for the earth. *Journal of Geophysical Research* 90: 2995–3007.
Christensen UR (2010) Dynamo scaling laws and applications to the planets. *Space Science Reviews* 152: 565–590.
Clauser C, et al. (1997) The thermal regime of the crystalline continental crust: Implications from the KTB. *Journal of Geophysical Research* 102: 18417–18441.
Cochran JR (1981) Simple models of diffuse extension and the pre-seafloor spreading development of the continental margin of the Northeastern Gulf of Aden. *Oceanologica Acta* 4(supplement): 155–165.
Cogley JG (1984) Continental margins and the extent and number of continents. *Reviews of Geophysics and Space Physics* 22: 101–122.
Cogné J-P and Humler E (2004) Temporal variation of oceanic spreading and crustal production rates during the last 180 My. *Earth and Planetary Science Letters* 227: 427–439.
Coltice N, Rolf T, Tackley PJ, and Labrosse S (2012) Dynamic causes of the relation between area and age of the ocean floor. *Science* 336: 335–338.
Condie KC (1998) Episodic continental growth and supercontinents: A mantle avalanche connection? *Earth and Planetary Science Letters* 163: 231–246.
Conrad CP and Hager BH (1999) The thermal evolution of an earth with strong subduction zones. *Geophysical Research Letters* 26: 3041–3044.
Courtney R and White R (1986) Anomalous heat flow and geoid across the Cape Verde Rise: Evidence of dynamic support from a thermal plume in the mantle. *Geophysical Journal of the Royal Astronomical Society* 87: 815–868.
Crane K and O'Connell S (1983) The distribution and implications of the heat flow from the Gregory rift in Kenya. *Tectonophysics* 94: 253–272.
Crosby AG and McKenzie D (2009) An analysis of young ocean depth, gravity and global residual topography. *Geophysical Journal International* 178: 1198–1219.
Crough ST (1983) Hotspot swells. *Annual Review of Earth and Planetary Sciences* 11: 165–193.
Cull JP (1991) Heat flow and regional geophysics in Australia. In: Cermak V and Rybach L (eds.) *Terrestrial Heat Flow and the Lithosphere Structure*, pp. 486–500. Berlin: Springer.
Davaille A and Jaupart C (1993) Transient high-Rayleigh-number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.
Davies GF (1980a) Thermal histories of convective Earth models and constraints on radiogenic heat production in the Earth. *Journal of Geophysical Research* 85: 2517–2530.
Davies GF (1980b) Review of oceanic and global heat flow estimates. *Reviews of Geophysics and Space Physics* 18: 718–722.
Davies GF (1988) Ocean bathymetry and mantle convection, 1. Large-scale flow and hotspots. *Journal of Geophysical Research* 93: 10467–10480.
Davies GF (1999) *Dynamic Earth: Plates, Plumes, and Mantle Convection*. Cambridge: Cambridge University Press.
Davies JH and Davies DR (2010) Earth's surface heat flux. *Solid Earth* 1: 5–24.
Davis EE and Elderfield H (2004) *Hydrogeology of the Oceanic Lithosphere*. Cambridge: Cambridge University Press.
Davis EE and Lister CRB (1974) Fundamentals of ridge crest topography. *Earth and Planetary Science Letters* 21: 405–413.
Davis EE, et al. (1999) Regional heat flow variations across the sedimented Juan de Fuca ridge eastern flank: Constraints on lithospheric cooling and lateral hydrothermal heat transport. *Journal of Geophysical Research* 104: 17675–17688.
de Koker N, Karki BB, and Stixrude LP (2013) Thermodynamics of the MgO–SiO_2 liquid system in Earth's lowermost mantle from first principles. *Earth and Planetary Science Letters* 361: 58–63.
de Koker N, Steinle-Neumann G, and Vlček V (2012) Electrical resistivity and thermal conductivity of liquid Fe alloys at high P and T, and heat flux in earth's core. *Proceedings of the National Academy of Sciences of the United States of America* 109: 4070–4073.
Deming D, Sass JH, and Lachenbruch AH (1992) and R. F. D. Rito1, Heat flow and subsurface temperature as evidence for basin-scale ground-water flow, North Slope of Alaska. *Geological Society of America Bulletin* 104: 528–542.

Donnelly KE, Goldstein SL, Langmuir CH, and Spiegelman M (2004) Origin of enriched ocean ridge basalts and implications for mantle dynamics. *Earth and Planetary Science Letters* 226: 347–366.

Duchkov AD (1991) Review of Siberian heat flow data. In: Cermak V and Rybach L (eds.) *Terrestrial Heat Flow and the Lithosphere Structure*, pp. 426–443. Berlin: Springer.

Dunn RA, Toomey DR, and Solomon SC (2000) Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9 30′N. *Journal of Geophysical Research* 105: 23537–23555.

Dye ST (2010) Geo-neutrinos and silicate earth enrichment of U and Th. *Earth and Planetary Science Letters* 297: 1–9.

Dye ST (2012) Geoneutrinos and the radioactive power of the Earth. *Reviews of Geophysics* 50: 3007.

Dye ST, Guillian E, Learned JG, et al. (2006) Earth radioactivity measurements with a deep ocean anti-neutrino observatory. *Earth, Moon, and Planets* 99: 241–252.

Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.

Elkins-Tanton LT, Parmentier EM, and Hess PC (2003) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteoritics and Planetary Science* 38: 1753–1771.

Engebretson DC, Cox A, and Gordon RG (1984) Relative motions between oceanic plates of the Pacific basin. *Journal of Geophysical Research* 89: 10291–10310.

England PC, Oxburg ER, and Richardson SW (1980) Heat refraction in and around granites in north-east England. *Geophysical Journal of the Royal Astronomical Society* 62: 439–455.

Enomoto S, Ohtani E, Inoue K, and Suzuki A (2007) Neutrino geophysics with KamLAND and future prospects. *Earth and Planetary Science Letters* 258: 147–159.

Erickson AJ, Von Herzen RP, Sclater JG, Girdler RW, Marshall BV, and Hyndman R (1975) Geothermal measurements in Deep-Sea drill holes. *Journal of Geophysical Research* 80: 2515–2528.

Fiorentini G, Lissia M, Mantovani F, and Vannucci R (2005) Geo-neutrinos: A new probe of Earth's interior [rapid communication]. *Earth and Planetary Science Letters* 238: 235–247.

Fiquet G, Auzende AL, Siebert J, et al. (2010) Melting of peridotite to 140 gigapascals. *Science* 329: 1516–1518.

Fitoussi C and Bourdon B (2012) Silicon isotope evidence against an enstatite chondrite Earth. *Science* 335: 1477–1480.

Flasar FM and Birch F (1973) Energetics of core formation: A correction. *Journal of Geophysical Research* 78: 6101–6103.

Fountain DM and Salisbury MH (1981) Exposed cross-sections through the continental crust: Implications for crustal structure, petrology, and evolution. *Earth and Planetary Science Letters* 56: 263–277.

Fountain DM, Salisbury MH, and Furlong KP (1987) Heat production and thermal conductivity of rocks from the Pikwitonei–Sachigo continental cross-section, central Manitoba: Implications for the thermal structure of Archean crust. *Canadian Journal of Earth Sciences* 24: 1583–1594.

Fourier JBJ (1824) Remarques générales sur les températures du globe terrestre et des espaces planétaires. *Annales des Chimie et des Physique* 27: 136–167.

Fournier RO (1989) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. *Annual Review of Earth and Planetary Sciences* 17: 13–53.

Francheteau J, Jaupart C, Jie SX, et al. (1984) High heat flow in southern Tibet. *Nature* 307: 32–36.

Golabek GJ, Schmeling H, and Tackley PJ (2008) Earth's core formation aided by flow channelling instabilities induced by iron diapirs. *Earth and Planetary Science Letters* 271: 24–33.

Gomi H, Ohta K, Hirose K, et al. (2013) The high conductivity of iron and thermal evolution of the earth's core. *Physics of the Earth and Planetary Interiors* 224: 88–103.

Gordon RG (2000) Diffuse oceanic plate boundaries: Strain rates, vertically averaged rheology, and comparisons with narrow plate boundaries and stable plate interiors. In: Richards MA, Gordon RG, and van der Hilst RD (eds.) *The History and Dynamics of Global Plate Motions*. *Geophysical Monograph Series*, vol. 121, pp. 143–159. Washington, DC: American Geophysical Union.

Goutorbe B, Drab L, Loubet N, and Lucazeau F (2007) Heat-flow revisited on the eastern Canadian Shield shelf. *Terra Nova* 19: 381–386.

Goutorbe B, Lucazeau F, and Bonneville A (2008a) The thermal regime of South African continental margins. *Earth and Planetary Science Letters* 267: 256–265.

Goutorbe B, Lucazeau F, and Bonneville A (2008b) The surface heat flow and the mantle contribution on the margins of Australia. *Geochemistry, Geophysics, Geosystems* 267: 256–265.

Green DH (1975) Genesis of Archean peridotitic magmas and constraints on Archean geothermal gradients and tectonics. *Geology* 3: 15–18.

Grigné C and Labrosse S (2001) Effects of continents on Earth cooling: Thermal blanketing and depletion in radioactive elements. *Geophysical Research Letters* 28: 2707–2710.

Grigné C, Labrosse S, and Tackley PJ (2005) Convective heat transfer as a function of wavelength: Implications for the cooling of the Earth. *Journal of Geophysical Research* 110: B03409. http://dx.doi.org/10.1029/2004JB003376.

Grigne C, Labrosse S, and Tackley PJ (2007) Convection under a lid of finite conductivity in wide aspect ratio models: Effect of continents on the wavelength of mantle flow. *Journal of Geophysical Research* 112(B08): 403–414.

Grocholski B, Catalli K, Shim S-H, and Prakapenka V (2012) Mineralogical effects on the detectability of the postperovskite boundary. *Proceedings of the National Academy of Sciences of the United States of America* 109: 2275–2279.

Grove TL and Parman SW (2004) Thermal evolution of the Earth as recorded by komatiites. *Earth and Planetary Science Letters* 219: 173–187.

Gubbins D and Roberts PH (1987) Magnetohydrodynamics of the Earth's core. In: Jacobs JA (ed.). *Geomagnetism*, vol. 2, pp. 1–183. London: Academic Press.

Gubbins D, Thomson CJ, and Whaler KA (1982) Stable regions in the Earth's liquid core. *Geophysical Journal of the Royal Astronomical Society* 68: 241–251.

Guillou L and Jaupart C (1995) On the effect of continents on mantle convection. *Journal of Geophysical Research* 100: 24217–24238.

Guillou L, Mareschal JC, Jaupart C, Gariépy C, Bienfait G, and Lapointe R (1994) Heat flow and gravity structure of the Abitibi belt, Superior Province, Canada. *Earth and Planetary Science Letters* 122: 447–460.

Guillou-Frottier L, Mareschal J-C, Jaupart C, Gariépy C, Lapointe R, and Bienfait G (1995) Heat flow variations in the Grenville Province, Canada. *Earth and Planetary Science Letters* 136: 447–460.

Gupta ML, Sharma SR, and Sundar A (1991) Heat flow and heat generation in the Archean Dharwar craton and implication for the southern Indian Shield geotherm. *Tectonophysics* 194: 107–122.

Harris RN and Chapman DS (2004) Deep-seated oceanic heat flow, heat deficits, and hydrothermal circulation. In: Davis EE and Elderfield H (eds.) *Hydrogeology of the Oceanic Lithosphere*, pp. 311–336. Cambridge: Cambridge University Press.

Hart SR and Zindler A (1986) In search of a bulk-Earth composition. *Chemical Geology* 57: 247–267.

Hasterok D (2013) A heat flow based cooling model for tectonic plates. *Earth and Planetary Science Letters* 361: 34–43.

Hasterok D, Chapman DS, and Davis EE (2011) Oceanic heat flow: Implications for global heat loss. *Earth and Planetary Science Letters* 311: 386–395.

Henry SG and Pollack HN (1988) Terrestrial heat flow above the Andean subduction zone in Bolivia and Peru. *Journal of Geophysical Research* 93: 15153–15162.

Henry P and Wang C-Y (1991) Modeling of fluid flow and pore pressure at the toe of Oregon and Barbados Accretionary Wedges. *Journal of Geophysical Research* 96: 20109–20130.

Hernlund JW and Labrosse S (2007) Geophysically consistent values of the perovskite to post-perovskite transition Clapeyron slope. *Geophysical Research Letters* 34: L05309.

Hernlund JW, Thomas C, and Tackley PJ (2005) Phase boundary double crossing and the structure of Earth's deep mantle. *Nature* 434: 882–886. http://dx.doi.org/10.1038/nature03472.

Herzberg C, Condie K, and Korenaga J (2010) Thermal history of the Earth and its petrological expression. *Earth and Planetary Science Letters* 292: 79–88.

Herzberg C and Zhang J (1996) Melting experiments on anhydrous peridotite KLB-1: Compositions of magmas in the upper mantle and transition zone. *Journal of Geophysical Research* 101: 8271–8296.

Hillier JK (2010) Subsidence of normal seafloor: Observations do indicate flattening. *Journal of Geophysical Research* 115: B03102.

Hillier JK and Watts AB (2005) Relationship between depth and age in the North Pacific Ocean. *Journal of Geophysical Research* 110: B02405.

Hirose K (2006) Postperovskite phase transition and its geophysical implications. *Reviews of Geophysics* 44: RG3001. http://dx.doi.org/10.1029/2005RG000186.

Hirose K, Gomi H, Ohta K, Labrosse S, and Hernlund JW (2011) The high conductivity of iron and thermal evolution of the Earth's core. *Mineralogical Magazine* 75: 1027.

Hirose K, Labrosse S, and Hernlund JW (2013) Composition and state of the core. *Annual Review of Earth and Planetary Sciences* 41: 657–691.

Hochmuth KA, Feilitzsch FV, Fields BD, et al. (2007) Probing the Earth's interior with a large-volume liquid scintillator detector. *Astroparticle Physics* 27: 21–29.

Hoffman PF (1997) Tectonic geology of North America. In: van der Pluijm B and Marshak S (eds.) *Earth Structure: An Introduction to Structural Geology and Tectonics*, pp. 459–464. New York: McGraw Hill.

Holmes A (1915) Radioactivity and the earth's thermal history: Part 1. The concentration of radioactive elements in the earth's crust. *Geological Magazine* 2: 60–71.

Howard LN (1964) Convection at high Rayleigh number. In: Gortler H (ed.) *Proceedings of the Eleventh International Congress of Applied Mechanics*, pp. 1109–1115. New York: Springer.
Hu S, He L, and Wang J (2000) Heat flow in the continental area of China: A new data set. *Earth and Planetary Science Letters* 179: 407–419.
Huismans R and Beaumont C (2011) Depth-dependent extension, two-stage breakup and cratonic underplating at rifted margins. *Nature* 473: 74–78.
Hulot G, Eymin C, Langlais B, Mandea M, and Olsen N (2002) Small-scale structure of the geodynamo inferred from Oersted and Magsat satellite data. *Nature* 416: 620–623.
Hulot G, Khoklov A, and LeMouel JL (1967) Uniqueness of mainly dipolar magnetic field recovered from directional data. *Geophysical Journal International* 129: 347–354.
Humler E, Langmuir C, and Daux V (1999) Depth versus age: New perspectives from the chemical compositions of ancient crust. *Earth and Planetary Science Letters* 173: 7–23.
Husson L, Pichon XLL, Henry P, Flotte N, and Rangin C (2008) Thermal regime of the NW shelf of the Gulf of Mexico, Part B: Heat flow. *Bulletin de la Société Géologique de France* 179: 139–145.
Hyndman RD (2010) The consequences of Canadian Cordillera thermal regime in recent tectonics and elevation: A review. *Canadian Journal of Earth Sciences* 47: 621–632.
Hyndman RD, Wang K, Yuan T, and Spence GD (1993) Tectonic sediment thickening, fluid expulsion, and the thermal regime of subduction zone accretionary prisms: The Cascadia margin off Vancouver Island. *Journal of Geophysical Research* 98: 21865–21876.
Ichikawa H, Labrosse S, and Kurita K (2010) Direct numerical simulation of an iron rain in the magma ocean. *Journal of Geophysical Research* 115: B01404.
Irifune T and Isshiki M (1998) Iron partitioning in a pyrolite mantle and the nature of the 410-km seismic discontinuity. *Nature* 392: 702–705.
Ito E and Katsura T (1989) A temperature profile of the mantle transition zone. *Geophysical Research Letters* 16: 425–428.
Jackson MG and Jellinek AM (2013) Major and trace element composition of the high $^{3}He/^{4}He$ mantle: Implications for the composition of a nonchonditic earth. *Geochemistry, Geophysics, Geosystems* 14: 2954–2976.
Jacobs JA (1961) Some aspects of the thermal history of the Earth. *Geophysical Journal of the Royal Astronomical Society* 4: 267–275.
Jacobs J and Allan DW (1956) The thermal history of the earth. *Nature* 177: 155–157.
Jarvis GT and Campbell IH (1983) Archean komatiites and geotherms – Solution to an apparent contradiction. *Geophysical Research Letters* 10: 1133–1136.
Jaupart C (1983) Horizontal heat transfer due to radioactivity contrasts: Causes and consequences of the linear heat flow-heat production relationship. *Geophysical Journal of the Royal Astronomical Society* 48: 411–435.
Jaupart C, Francheteau J, and Shen X-J (1985) On the thermal structure of the southern Tibetan crust. *Geophysical Journal of the Royal Astronomical Society* 81: 131–155.
Jaupart C, Mann JR, and Simmons G (1982) A detailed study of the distribution of heat flow and radioactivity in New Hampshire (U.S.A.). *Earth and Planetary Science Letters* 59: 267–287.
Jaupart C and Mareschal JC (1999) The thermal structure and thickness of continental roots. *Lithos* 48: 93–114.
Jaupart C and Mareschal JC (2013) Constraints on crustal heat production from heat flow data. In: Rudnick RL (ed.) *Treatise on Geochemistry. The Crust*, 2nd edn., vol. 3, pp. 65–84. New York: Pergamon.
Javoy M (1995) The integral enstatite chondrite model of the earth. *Geophysical Research Letters* 22: 2219–2222.
Javoy M (1999) Chemical Earth models. *Comptes Rendus de l'Acadêmie des Sciences, Paris* 329: 537–555.
Javoy M, et al. (2010) The chemical composition of the Earth: Enstatite chondrite models. *Earth and Planetary Science Letters* 293: 259–268.
Jeffreys S (1962) *The Earth: Its Origin, History, and Physical Constitution*, 5th edn. Cambridge: Cambridge University Press.
Joeleht TH and Kukkonen IT (1998) Thermal properties of granulite facies rocks in the Precambrian basement of Finland and Estonia. *Tectonophysics* 291: 195–203.
Joly J (1909) *Radioactivity and Geology*. Dublin: Dublin University Press.
Jones MQW (1988) Heat flow in the Witwatersrand Basin and environs and its significance for the South African Shield geotherm and lithosphere thickness. *Journal of Geophysical Research* 93: 3243–3260.
Kaminski E and Javoy M (2013) A two-stage scenario for the formation of the earth's mantle and core. *Earth and Planetary Science Letters* 365: 97–107.
Kamland Collaboration (2011) Partial radiogenic heat model for Earth revealed by geoneutrino measurements. *Nature Geosciences* 4: 647–651.
Kamland Collaboration (Gando, A., 45 Collaborators) (2013) Reactor on-off antineutrino measurement with KamLAND. *Physical Review D* 88: 033001.
Katsura T, et al. (2003) Post-spinel transition in $MgSiO_4$ determined by in-situ X-ray diffractometry. *Physics of the Earth and Planetary Interiors* 136: 11–24.
Katsura T, et al. (2004) Olivine–Wadsleyite transition in the system (Mg, Fe)SiO_4. *Journal of Geophysical Research* 109: B02209.
Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283: 1881–1884.
Ketcham RA (1996) Distribution of heat-producing elements in the upper and middle crust of southern and west central Arizona: Evidence from the core complexes. *Journal of Geophysical Research* 101(13): 13611–13632.
King SD and Anderson DL (1998) Edge-driven convection. *Earth and Planetary Science Letters* 160: 289–296.
Kinzler RJ and Grove TL (1992) Primary magmas of mid-ocean ridge basalts. 2. Applications. *Journal of Geophysical Research* 97: 6907–6926.
Klein EM and Langmuir CH (1987) Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *Journal of Geophysical Research* 92: 8089–8115.
Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.
Korenaga J (2003) Energetics of mantle convection and the fate of fossil heat. *Geophysical Research Letters* 30: 20. http://dx.doi.org/10.1029/2003GL016982.
Korenaga J (2006) Archean geodynamics and the thermal evolution of the earth. In: Benn K, Mareschal J, and Condie KC (eds.) *Archean Geodynamics and Environments*. *Geophysical Monograph Series*, vol. 164, pp. 7–32. Washington, DC: American Geophysical Union.
Korenaga T and Korenaga J (2008) Subsidence of normal oceanic lithosphere, apparent thermal expansivity, and sea floor flattening. *Earth and Planetary Science Letters* 268: 41–51.
Kremenentsky AA, Milanovsky SY, and Ovchinnikov LN (1989) A heat generation model for the continental crust based on deep drilling in the Baltic shield. *Tectonophysics* 159: 231–246.
Kukkonen IT and Peltonen P (1999) Xenolith-controlled geotherm for the central Fennoscandian shield: Implications for lithosphere–asthenosphere relations. *Tectonophysics* 304: 301–315.
Labrosse S (2002) Hotspots, mantle plumes and core heat loss. *Earth and Planetary Science Letters* 199: 147–156.
Labrosse S (2003) Thermal and magnetic evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 140: 127–143.
Labrosse S (2005a) Core-mantle boundary, heat flow across. In: Gubbins D and Herrero-Bervera E (eds.) *Encyclopedia of Geomagnetism and Paleomagnetism*, pp. 127–130. Dordrecht, The Netherlands: Springer.
Labrosse S (2005b) Geodynamo, energy sources. In: Gubbins D and Herrero-Bervera E (eds.) *Encyclopedia of Geomagnetism and Paleomagnetism*, pp. 300–302. Dordrecht, The Netherlands: Springer.
Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of Earth's mantle. *Nature* 450: 866–869.
Labrosse S and Jaupart C (2007) Thermal evolution of the Earth: Secular changes and fluctuations of plate characteristics. *Earth and Planetary Science Letters* 260: 465–481.
Labrosse S, Poirier J-P, and Le Mouël J-L (1997) On cooling of the Earth's core. *Physics of the Earth and Planetary Interiors* 99: 1–17.
Lachenbruch AH (1970) Crustal temperature and heat production: Implications of the linear heat flow heat production relationship. *Journal of Geophysical Research* 73: 3292–3300.
Lachenbruch AH and Sass JH (1978) Models of an extending lithosphere and heat flow in the Basin and Range province. *Memoirs of the Geological Society of America* 152: 209–258.
Lallemand S, Heuret A, and Boutelier D (2005) On the relationships between slab dip, back-arc stress, upper plate absolute motion, and crustal nature in subduction zones. *Geochemistry, Geophysics, Geosystems* 6: 1–18. http://dx.doi.org/10.1029/2005GC000917.
Lambeck K (1977) Tidal dissipation in the oceans: Astronomical, geophysical, and oceanographic consequences. *Philosophical Transactions of the Royal Society of London, Series A* 287: 545–594.
Langmuir CH, Goldstein SL, Donnelly K, and Su YK (2005) Origins of enriched and depleted mantle reservoirs. *EOS Transactions*. *Fall Meeting Supplement*, vol. 86, p. 1. Washington, DC: American Geophysical Union, Abstract V23D-02.
Lapwood ER (1952) The effect of contraction in the cooling of a gravitating sphere, with special reference to the Earth. *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement* 6: 402–407.

Latil-Brun MV and Lucazeau F (1988) Subsidence, extension and thermal history of the West African margin in Senegal. *Earth and Planetary Science Letters* 90: 204–220.

Lay T, Hernlund J, Garnero EJ, and Thorne MS (2006) A post-perovskite lens and D″ heat flux beneath the central Pacific. *Science* 314: 1272. http://dx.doi.org/10.1126/science.1133280.

Lenardic A, Moresi L-N, Jellinek A, and Manga M (2005) Continental insulation, mantle cooling, and the surface area of oceans and continents. *Earth and Planetary Science Letters* 234: 317–333.

Leng W and Zhong S (2010) Constraints on viscous dissipation of plate bending from compressible mantle convection. *Earth and Planetary Science Letters* 297: 154–164.

Lévy F, Jaupart C, Mareschal J, Bienfait G, and Limare A (2010) Low heat flux and large variations of lithospheric thickness in the Canadian Shield. *Journal of Geophysical Research* 115: 6404.

Lewis TJ, Hyndman RD, and Fluck P (2003) Heat flow, heat generation and crustal temperatures in the northern Canadian Cordillera: Thermal control on tectonics. *Journal of Geophysical Research* 108: 2316.

Lister CRB (1977) Estimators for heat flow and deep rock properties based on boundary layer theory. *Tectonophysics* 41: 157–171.

Lister JR (2003) Expressions for the dissipation driven by convection in the Earth's core. *Physics of the Earth and Planetary Interiors* 140: 145–158.

Lister JR and Buffett BA (1995) The strength and efficiency of the thermal and compositional convection in the geodynamo. *Physics of the Earth and Planetary Interiors* 91: 17–30.

Lister JR and Buffett BA (1998) Stratification of the outer core at the core–mantle boundary. *Physics of the Earth and Planetary Interiors* 105: 5–19.

Lister CRB, Sclater JG, Nagihara S, Davis EE, and Villinger H (1990) Heat flow maintained in ocean basins of great age – Investigations in the north-equatorial West Pacific. *Geophysical Journal International* 102: 603–630.

Litasov K and Ohtani E (2002) Phase relations and melt compositions in CMAS–pyrolite–H_2O system up to 25 GPa. *Physics of the Earth and Planetary Interiors* 134: 105–127.

Louden KE, Sibuet J-C, and Harmegnies F (1997) Variations in heat flow across the ocean–continent transition in the Iberia abyssal plain. *Earth and Planetary Science Letters* 151: 233–254.

Lucazeau F, Bonneville A, Escartin J, et al. (2006) Heat flow variations on a slowly accreting ridge: Constraints on the hydrothermal and conductive cooling for the Lucky Strike segment (Mid-Atlantic Ridge, 37°N). *Geochemistry, Geophysics, Geosystems* 7: 7011.

Lucazeau F, Brigaud F, and Bouroullec JL (2004) High-resolution heat flow density in the lower Congo basin. *Geochemistry, Geophysics, Geosystems* 5: 3001.

Lucazeau F, Leroy S, Rolandone F, et al. (2010) Heat-flow and hydrothermal circulation at the ocean–continent transition of the eastern Gulf of Aden. *Earth and Planetary Science Letters* 295: 554–570.

Lucazeau F, et al. (2008) Persistent thermal activity at the Eastern Gulf of Aden after continental break-up. *Nature Geoscience* 1: 854–858.

Lyubetskaya T and Korenaga J (2007) Chemical composition of Earth's primitive mantle and its variance: 1. Method and results. *Journal of Geophysical Research* 112: 3211.

MacDonald GJF (1959) Calculations on the thermal history of the Earth. *Journal of Geophysical Research* 64: 1967–2000.

Marcaillou B, Spence G, Collot J-Y, and Wang K (2006) Thermal regime from bottom simulating reflectors along the north Ecuador-south Colombia margin: Relation to margin segmentation and great subduction earthquakes. *Journal of Geophysical Research* 111: 12407.

Mareschal JC and Jaupart C (2004) Variations of surface heat flow and lithospheric thermal structure beneath the North American craton. *Earth and Planetary Science Letters* 223: 65–77.

Mareschal JC, Jaupart C, Cheng LZ, et al. (1999) Heat flow in the Trans-Hudson Orogen of the Canadian Shield: Implications for Proterozoic continental growth. *Journal of Geophysical Research* 104: 29007–29024.

Mareschal JC, Jaupart C, Gariépy C, et al. (2000a) Heat flow and deep thermal structure near the southeastern edge of the Canadian Shield. *Canadian Journal of Earth Sciences* 37: 399–414.

Mareschal J-C, Jaupart C, Phaneuf C, and Perry C (2012) Geoneutrinos and the energy budget of the Earth. *Journal of Geodynamics* 54: 43–54.

Mareschal JC, Jaupart C, Rolandone F, et al. (2005) Heat flow, thermal regime, and rheology of the lithosphere in the Trans-Hudson Orogen. *Canadian Journal of Earth Sciences* 42: 517–532.

Mareschal JC, Poirier A, Rolandone F, et al. (2000b) Low mantle heat flow at the edge of the North American Continent, Voisey Bay, Labrador. *Geophysical Research Letters* 27: 823–826.

Martínez F and Cochran JR (1989) Geothermal measurements in the northern Red Sea: Implications for lithospheric thermal structure and mode of extension during continental rifting. *Journal of Geophysical Research* 94: 12239–12265.

McDonald GJF (1959) Calculations on the thermal history of the earth. *Journal of Geophysical Research* 64: 1967–2000.

McDonough WF and Sun SS (1995) The composition of the earth. *Chemical Geology* 120: 223–253.

McKenna TE and Sharp JM Jr. (1998) Radiogenic heat production in sedimentary rocks of the Gulf of Mexico basin, south Texas. *American Association of Petroleum Geologists Bulletin* 82: 484–496.

McKenzie D (1978) Some remarks on the development of sedimentary basins. *Earth and Planetary Science Letters* 40: 25–32.

McKenzie D and Bickle M (1988) The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology* 29: 625–679.

McKenzie D, Jackson J, and Priestley K (2005) Thermal structure of oceanic and continental lithosphere. *Earth and Planetary Science Letters* 233: 337–349.

McKenzie D and Richter FM (1981) Parameterized thermal convection in a layered region and the thermal history of the Earth. *Journal of Geophysical Research* 86: 11667–11680.

McKenzie DP and Weiss N (1975) Speculations on the thermal and tectonic history of the earth. *Geophysical Journal* 42: 131–174.

McLennan J, Hulme T, and Singh SC (2005) Cooling of the lower oceanic crust. *Geology* 33: 357.

Melosh HJ and Ivanov BA (1999) Impact crater collapse. *Annual Review of Earth and Planetary Sciences* 27: 385–415.

Mittelstaedt E, Escartín J, Gracias N, et al. (2012) Quantifying diffuse and discrete venting at the Tour Eiffel vent site, Lucky Strike hydrothermal field. *Geochemistry, Geophysics, Geosystems* 13: 4008.

Mojzsis SJ, Harrison TM, and Pidgeon RT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409: 178–181.

Molnar P and England P (1990) Temperatures, heat flux, and frictional stress near major thrust faults. *Journal of Geophysical Research* 95: 4833–4856.

Monteux J, Ricard YR, Coltice N, Dubuffet F, and Ulvrová M (2009) A model of metal–silicate separation on growing planets. *Earth and Planetary Science Letters* 287: 353–362.

Morgan P (1983) Constraints on rift thermal processes from heat flow and uplift. *Tectonophysics* 94: 277–298.

Müller RD, Roest WR, Royer J-Y, Gahagan LM, and Sclater JG (1997) Digital isochrons of the world's ocean floor. *Journal of Geophysical Research* 102: 3211–3214.

Müller RD, Sdrolias M, Gaina C, and Roest WR (2008) Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochemistry, Geophysics, Geosystems* 9: 4006.

Munk WH and MacDonald GJF (1960) *The Rotation of the Earth*. Cambridge: Cambridge University Press.

Murakami M, Hirose K, Kawamura K, Sata N, and Ohishi Y (2004) Post-perovskite phase transition in $MgSiO_3$. *Science* 304: 855–858.

Nagihara S and Opre K (2005) Jones, Geothermal heat flow in the northeast margin of the Gulf of Mexico. *American Association of Petroleum Geologists Bulletin* 89: 821–831.

Newsom HE and Jones JH (eds.) (1990) *Origin of the Earth*. Oxford: Oxford University Press.

Nicolaysen LO, Hart RJ, and Gale NH (1981) The Vredefort radioelement profile extended to supracrustal strata at Carletonville, with implications for continental heat flow. *Journal of Geophysical Research* 86: 10653–10662.

Nisbet EG, Cheadle MJ, Arndt NT, and Bickle MJ (1995) Constraining the potential temperature of the Archean mantle: A review of the evidence from komatiites. *Lithos* 30: 291–307.

Nomura R, Ozawa H, Tateno S, et al. (2011) Spin crossover and iron-rich silicate melt in the Earth's deep mantle. *Nature* 473: 199–202.

Nyblade AA (1997) Heat flow across the East African Plateau. *Geophysical Research Letters* 24: 2083–2086.

Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of $MgSiO_3$ in Earth's D″ layer. *Nature* 430: 445–448.

Ogawa M, Schubert G, and Zebib A (1991) Numerical simulations of three-dimensional thermal convection in a fluid with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 233: 299–328.

Olson P (1987) A comparison of heat transfer laws for mantle convection at very high Rayleigh numbers. *Physics of the Earth and Planetary Interiors* 48: 153–160.

Olson P (1990) Hot spots, swells and mantle plumes. In: *Magma Transport and Storage*, pp. 33–51. New York: Wiley.

Palme H and O'Neill HSC (2003) Cosmochemical estimates of mantle composition. In: Carlson RW (ed.) *Mantle and Core*. *Treatise on Geochemistry*, vol. 2. Amsterdam: Elsevier.
Pari G and Peltier WR (2000) Subcontinental mantle dynamics: A further analysis based on the joint constraints of dynamic surface topography and free-air gravity. *Journal of Geophysical Research* 105: 5635–5662.
Parsons B (1982) Causes and consequences of the relation between area and age of the ocean floor. *Journal of Geophysical Research* 87: 289–302.
Parsons B and Sclater JG (1977) An analysis of the variation of ocean floor bathymetry and heat flow with age. *Journal of Geophysical Research* 82: 803–827.
Peltier WR and Jarvis GT (1982) Whole mantle convection and the thermal evolution of the earth. *Physics of the Earth and Planetary Interiors* 29: 281–304.
Perry HKC, Jaupart C, Mareschal JC, and Bienfait G (2006) Crustal heat production in the Superior Province, Canadian Shield, and in North America inferred from heat flow data. *Journal of Geophysical Research* 111: B07301.
Perry HKC, Mareschal J, and Jaupart C (2009) Enhanced crustal geo-neutrino production near the Sudbury Neutrino Observatory, Ontario, Canada. *Earth and Planetary Science Letters* 288: 301–308.
Pinet C and Jaupart C (1987) The vertical distribution of radiogenic heat production in the Precambrian crust of Norway and Sweden: Geothermal implications. *Geophysical Research Letters* 14: 260–263.
Pinet C, Jaupart C, Mareschal J-C, Gariepy C, Bienfait G, and Lapointe R (1991) Heat flow and structure of the lithosphere in the eastern Canadian shield. *Journal of Geophysical Research* 96: 19941–19963.
Poirier J-P (2000) *Introduction to the Physics of the Earth's Interior*. 2nd edn. Cambridge: Cambridge University Press.
Pollack HN and Chapman DS (1977a) On the regional variations of heat flow, geotherms, and lithospheric thickness. *Tectonophysics* 38: 279–296.
Pollack HN and Chapman DS (1977b) Mantle heat flow. *Earth and Planetary Science Letters* 34: 174–184.
Pollack HN, Hurter SJ, and Johnston JR (1993) Heat flow from the earth's interior: Analysis of the global data set. *Reviews of Geophysics* 31: 267–280.
Poort J and Klerkx J (2004) Absence of a regional surface thermal high in the Baikal rift: New insights from detailed contouring of heat flow anomalies. *Tectonophysics* 383: 217–241.
Pozzo M, Davies CJ, Gubbins D, and Alfè D (2012) Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* 485: 355–358.
Putirka KD, Perfit M, Ryerson FJ, and Jackson MG (2007) Ambient and excess mantle temperatures, olivine thermometry, and active vs passive upwelling. *Chemical Geology* 241: 177–206.
Raghavan RS, Schoenert S, Enomoto S, Shirai J, Suekane F, and Suzuki A (1998) Measuring the global radioactivity in the Earth by multidetector antineutrino spectroscopy. *Physical Review Letters* 80: 635–638.
Ramondec P, Germanovich L, Damm KV, and Lowell R (2006) The first measurements of hydrothermal heat output at 9°50′N, East Pacific Rise. *Earth and Planetary Science Letters* 245: 487–497.
Rao YH, Subrahmanyam C, Sharma SR, Rastogi AA, and Deka B (2001) Estimates of geothermal gradients and heat flow from BSRs along the Western Continental Margin of India. *Geophysical Research Letters* 28: 355–358.
Ray RD, Eanes RJ, and Chao BF (1996) Detection of tidal dissipation in the solid earth by satellite tracking and altimetry. *Nature* 381: 595–597.
Reiter M and Jessop AM (1985) Estimates of terrestrial heat flow in offshore eastern Canada. *Canadian Journal of Earth Sciences* 22: 1503–1517.
Ribe NM and Christensen UR (1994) Three-dimensional modeling of plume–lithosphere interaction. *Journal of Geophysical Research* 99: 669–682.
Ricard Y, Šrámek O, and Dubuffet F (2009) A multi-phase model of runaway core–mantle segregation in planetary embryos. *Earth and Planetary Science Letters* 284: 144–150.
Ringwood AE (1962) A model for the upper mantle. *Journal of Geophysical Research* 67: 857–867.
Rolandone F, Jaupart C, Mareschal JC, et al. (2002) Surface heat flow, crustal temperatures and mantle heat flow in the Proterozoic Trans-Hudson Orogen, Canadian Shield. *Journal of Geophysical Research* 107: 2314. http://dx.doi.org/10.1029/2001JB000698.
Rothschild CG, Chen MC, and Calaprice FP (1998) Antineutrino geophysics with liquid scintillator detectors. *Geophysical Research Letters* 25: 1083–1086.
Rowley DB (2002) Rate of plate creation and destruction: 180 Ma to present. *Geological Society of America Bulletin* 114: 927–933.
Roy RF, Decker ER, Blackwell DD, and Birch F (1968) Heat flow in the United States. *Journal of Geophysical Research* 73: 5207–5221.
Roy S and Rao RUM (2000) Heat flow in the Indian shield. *Journal of Geophysical Research* 105: 25587–25604.
Royer J, Müller R, Gahagan L, et al. (1992) A global isochron chart. *Technical Report 117*, University of Texas Institute for Geophysics, Austin.
Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal–silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.
Rudnick RL and Fountain DM (1995) Nature and composition of the continental crust: A lower crustal perspective. *Reviews of Geophysics* 33: 267–310.
Rudnick RL and Gao S (2003) Composition of the continental crust. In: Rudnick RL (ed.) *Treatise on Geochemistry*. *The Crust*, vol. 3, pp. 1–64. New York: Pergamon.
Rudnick RL, McDonough WF, and O'Connell RJ (1998) Thermal structure, thickness and composition of continental lithosphere. *Chemical Geology* 145: 395–411.
Rudnick RL and Nyblade AA (1999) The thickness of Archean lithosphere: Constraints from xenolith thermobarometry and surface heat flow. In: Fei Y, Bertka CM, and Mysen BO (eds.) *Mantle Petrology: Field Observations and High Pressure Experimentation: A Tribute to Francis R. (Joe) Boyd*, pp. 3–11. Geochemical Society.
Rushmer T, Minarik WG, and Taylor GJ (2000) Physical processes of core formation. In: Canup RM and Drake K (eds.) *Origin of the Earth and Moon*, pp. 227–243. Tucson, AZ: University of Arizona Press.
Russell JK, Dipple GM, and Kopylova MG (2001) Heat production and heat flow in the mantle lithosphere, Slave craton, Canada. *Physics of the Earth and Planetary Interiors* 123: 27–44.
Rybach L (1988) Determination of heat production rate. In: Haenel R, Rybach L, and Stegena L (eds.) *Handbook of Terrestrial Heat-Flow Density Determination*, pp. 125–142. Dordrecht: Kluwer.
Schubert G, Stevenson D, and Cassen P (1980) Whole planet cooling and the radiogenic heat source contents of the Earth and Moon. *Journal of Geophysical Research* 85: 2531–2538.
Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.
Schubert G and Young RE (1976) Cooling the Earth by whole mantle subsolidus convection: A constraint on the viscosity of the lower mantle. *Tectonophysics* 35: 201–214.
Sclater JG and Crowe J (1979) A heat flow survey at anomaly 13 on the Reykjanes Ridge: A critical test of the relation between heat flow and age. *Journal of Geophysical Research* 84: 1593–1602.
Sclater JG and Francheteau J (1970) The implications of terrestrial heat flow observations on current tectonic and geochemical models of the crust and upper mantle of the earth. *Geophysical Journal of the Royal Astronomical Society* 20: 509–542.
Sclater JG, Jaupart C, and Galson D (1980) The heat flow through oceanic and continental crust and the heat loss of the Earth. *Reviews of Geophysics and Space Physics* 18: 269–312.
Seton M, Gaina C, Muller RD, and Heine C (2009) Mid-Cretaceous seafloor spreading pulse: Fact or fiction? *Geology* 37: 687–690.
Sharpe HN and Peltier WR (1978) Parameterized mantle convection and the Earth's thermal history. *Geophysical Research Letters* 5: 737–740.
Singh SC, Minshull TA, and Spence GD (1993) Velocity structure of a gas hydrate reflector. *Science* 260: 204–207.
Sleep NH (1971) Thermal effects of the formation of Atlantic continental margins by continental break up. *Geophysical Journal of the Royal Astronomical Society* 24: 325–350.
Sleep NH (1979) Thermal history and degassing of the earth: Some simple calculations. *Journal of Geology* 87: 671–686.
Sleep NH (1990) Hotspots and mantle plumes: Some phenomenology. *Journal of Geophysical Research* 95: 6715–6736.
Sleep NH (2000) Evolution of the mode of convection within terrestrial planets. *Journal of Geophysical Research* 105: 17563–17578.
Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM, Righter K, and 69 Collaborating Authors (eds.) *Origin of the Earth and Moon*, pp. 323–338. Tucson, AZ: University of Arizona Press.
Solomatov VS and Moresi LN (1997) Three regimes of mantle convection with non-Newtonian viscosity and stagnant lid convection on the terrestrial planets. *Geophysical Research Letters* 24: 1907–1910.
Sotin C and Labrosse S (1999) Three-dimensional thermal convection of an isoviscous, infinite Prandtl number fluid heated from within and from below: Applications to heat transfer in planetary mantles. *Physics of the Earth and Planetary Interiors* 112: 171–190.
Spinelli GA and Harris RN (2011) Effects of the legacy of axial cooling on partitioning of hydrothermal heat extraction from oceanic lithosphere. *Journal of Geophysical Research* 116: 9102.

Šrámek O, Milelli L, Ricard Y, and Labrosse S (2012) Thermal evolution and differentiation of planetesimals and planetary embryos. *Icarus* 217: 339–354.
Šrámek O, Ricard Y, and Dubuffet F (2010) A multiphase model of core formation. *Geophysical Journal International* 181: 198–220.
Stacey FD and Davis PM (2008) *Physics of the Earth*. 4th edn. Cambridge: Cambridge University Press.
Stein CA and Stein S (1992) A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* 359: 123–129.
Stevenson DJ (1989) Formation and early evolution of the Earth. In: Peltier W (ed.) *Mantle Convection*, pp. 817–873. New York: Gordon and Breach.
Stevenson DJ (1990) Fluid dynamics of core formation. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 231–249. New York: Oxford University Press.
Stevenson DJ (2003) Mission to earth's core – A modest proposal. *Nature* 423: 239–240.
Stixrude L (1997) Structure and sharpness of phase transitions and mantle discontinuities. *Journal of Geophysical Research* 102: 14835–14852.
Strutt RJ (1906) On the distribution of radium in the earth's crust and on internal heat. *Proceedings of the Royal Society of London, Series A* 77: 472–485.
Su YJ (2000) *Mid-Ocean Ridge Basalt Trace Element Systematics: Constraints from Database Management, ICPMS Analysis, Global Data Compilation and Petrologic Modeling*. Unpublished PhD Thesis, Columbia University, New York, No. 1, 569 pp.
Swanberg CA, Chessman MD, Simmons G, Smithson SB, Grnlie G, and Heier KS (1974) Heat-flow heat-generation studies in Norway. *Tectonophysics* 23: 31–48.
Tackley PJ (2000) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations, 1. Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* 1: 1021.
Tateno S, Hirose K, Sata N, and Ohishi Y (2009) Determination of post-perovskite phase transition boundary up to 4400 K and implications for thermal structure in D″ layer. *Earth and Planetary Science Letters* 277: 130–136.
Thomas CW, Liu Q, Agee CB, Asimow PD, and Lange RA (2012) Multi-technique equation of state for Fe_2SiO_4 melt and the density of Fe-bearing silicate melts from 0 to 161 GPa. *Journal of Geophysical Research* 117: B10206.
Thomson W (1864) On the secular cooling of the earth. *Transactions of the Royal Society of Edinburgh* 23: 295–311.
Turcotte DL and Oxburgh ER (1967) Finite amplitude convective cells and continental drift. *Journal of Fluid Mechanics* 28: 29–42.
Ulvrová M, Coltice N, Ricard Y, Labrosse S, Dubuffet F, and Sramek O (2011) Compositional and thermal equilibration of particles, drops and diapirs in geophysical flows. *Geochemistry, Geophysics, Geosystems* 12: 1–11.
Unternehr P, Peron-Pinvidic G, Manatschal G, and Sutra E (2010) Hyper-extended crust in the South-Atlantic: In search of a model. *Petroleum Geoscience* 16: 207–215.
Urey HC (1964) A review of atomic abundances in chondrites and the origin of meteorites. *Reviews of Geophysics* 2: 1–34.
Vasseur G and Singh RN (1986) Effects of random horizontal variations in radiogenic heat source distribution on its relationship with heat flow. *Journal of Geophysical Research* 91: 10397–10404.
Villinger H, Grevemeyer I, Kaul N, Hauschild J, and Pfender M (2002) Hydrothermal heat flux through aged oceanic crust: Where does the heat escape? *Earth and Planetary Science Letters* 202: 159–170.
Vočadlo L, Alfè D, Gillan MJ, and Price GD (2003) The properties of iron under core conditions from first principles calculations. *Physics of the Earth and Planetary Interiors* 140: 101–125.
Vogt PR and Ostenso NA (1967) Steady state crustal spreading. *Nature* 215: 810–817.
Vosteen HD, Rath V, Clauser C, and Lammerer B (2003) The thermal regime of the Eastern Alps from inverse analyses along the TRANSALP profile. *Physics and Chemistry of the Earth* 28: 393–405.
Wasserburg G, McDonald GJF, Hoyle F, and Fowler WA (1964) Relative contributions of uranium, thorium, and potassium to heat production in the earth. *Science* 143: 465–467.
Wasson JT and Kalleymen GW (1988) Composition of chondrites. *Philosophical Transactions of the Royal Society of London* 325: 535–544.
Wetherill GW (1990) Formation of the Earth. *Annual Review of Earth and Planetary Sciences* 18: 205–256.
Whitmarsh RB, Manatschal G, and Minshull TA (2001) Evolution of magma-poor continental margins from rifting to seafloor spreading. *Nature* 413: 150–154.
Williams Q and Garnero EJ (1996) Seismic evidence for partial melt at the base of the Earth's mantle. *Science* 273: 1528.
Williams DL, Herzen RP, Sclater JG, and Anderson RN (1974) The Galapagos Spreading Centre: Lithospheric cooling and hydrothermal circulation. *Geophysical Journal of the Royal Astronomical Society* 38: 587–608.
Williams DL and von Herzen RP (1974) Heat loss from the earth: New estimate. *Geology* 2: 327–328.
Workman RK and Hart SR (2005) Major and trace element composition of the depleted MORB mantle (DMM). *Earth and Planetary Science Letters* 231: 53–72.
Zerr A, Diegeler A, and Boehler R (1998) Solidus of Earth's deep mantle. *Nature* 281: 243–246.
Zschau J (1986) Constraints from the Chandler wobble period. In: Anderson AJ and Cazenave A (eds.) *Space Geodesy and Geodynamics*, pp. 315–344. New York: Academic Press.

7.07 The Generation of Plate Tectonics from Mantle Dynamics

D Bercovici, Yale University, New Haven, CT, USA
PJ Tackley, Institut für Geophysik, ETH Zürich, Switzerland
Y Ricard, Ecole Normale Supérieure de Lyon, Université de Lyon-1, CNRS, Lyon, France

© 2015 Elsevier B.V. All rights reserved.

7.07.1 Introduction 272
7.07.2 Plate Tectonics 273
7.07.2.1 Present-Day Plate Motions 273
7.07.2.1.1 Plate kinematics 273
7.07.2.1.2 Plate forces 275
7.07.2.1.3 Plate age and size distributions 275
7.07.2.2 Origin and History of Plate Tectonics 275
7.07.2.3 Planetary Tectonics 276
7.07.3 Basic Convection 276
7.07.3.1 Convective Instability and the Rayleigh Number 276
7.07.3.2 Vertical Structure of Simple Convection 277
7.07.3.2.1 Symmetry 277
7.07.3.2.2 Thermal boundary layers 277
7.07.3.2.3 The size of convection cells 278
7.07.3.2.4 Thermal boundary layer forces 278
7.07.3.3 Patterns of Convection 279
7.07.3.4 Influence of Internal Heating 280
7.07.3.5 Influence of Temperature-Dependent Viscosity 281
7.07.3.6 A Brief Note on the Influence of Sphericity 283
7.07.3.7 Poloidal and Toroidal Flow 284
7.07.4 Where Does Basic Convection Theory Succeed in Explaining Plate Tectonics? 285
7.07.4.1 Convective Forces and Plate Driving Forces 285
7.07.4.1.1 Slab pull and convective downwellings 285
7.07.4.1.2 Ridge push is a convective lateral pressure gradient 287
7.07.4.2 Structure of Ocean Basins 287
7.07.4.3 Slablike Downwellings Are Characteristic of 3-D Convection 288
7.07.4.3.1 Cylindrical upwellings and mantle plumes are also characteristic of 3-D convection 288
7.07.4.4 Slab Heat Flux and Plate Velocities 289
7.07.4.5 Mantle Heterogeneity and the History of Subduction 289
7.07.4.6 Temperature-Dependent Viscosity, Internal Resistive Boundaries, and the Aspect Ratio of Convection Cells 290
7.07.5 Where Does Basic Convection Theory Fail in Explaining Plate Tectonics? 290
7.07.5.1 Plate Forces Not Well Explained by Basic Convection 290
7.07.5.2 Structure of Ocean Basins (Deviations from the $\sqrt{\text{age}}$ Law) 291
7.07.5.3 Changes in Plate Motion 291
7.07.5.4 Platelike Strength Distributions: Weak Boundaries and Strong Interiors 292
7.07.5.4.1 Do plates form because the lithosphere breaks? 292
7.07.5.4.2 Plateness 292
7.07.5.5 Convergent Margins and Subduction Zones 294
7.07.5.5.1 Subduction initiation 294
7.07.5.5.2 Asymmetry of subduction 294
7.07.5.5.3 Age of subduction zones 294
7.07.5.5.4 Melting, crust and depleted lithosphere buoyancy, and the eclogite transition 295
7.07.5.5.5 Trench-parallel flow and rollback 296
7.07.5.6 Divergent Margins: Ridges and Narrow, Passive Upwellings 296
7.07.5.7 Strike-Slip Margins: Generation of Toroidal Motion 297
7.07.5.7.1 Imposing plates or faults 297
7.07.5.7.2 Why is variable viscosity necessary for toroidal motion? 298
7.07.6 Plate Generation Physics 298
7.07.6.1 Non-Newtonian Mechanics 298
7.07.6.1.1 What variable viscosities give both plateness and toroidal flow? 298
7.07.6.1.2 What is the purpose of toroidal motion? 299
7.07.6.1.3 Instantaneous rheologies and plate boundary history 299

7.07.6.2	Dynamic Weakening and Damage Physics	300
7.07.6.2.1	Thermal self-weakening	300
7.07.6.2.2	Water and void weakening	302
7.07.6.2.3	Grain-size weakening	304
7.07.6.2.4	Summary thoughts on weakening and damage	306
7.07.7	**Plate Generation on Earth and Other Planets**	306
7.07.7.1	Origin of Plate Tectonics and Archean Tectonics	306
7.07.7.2	Plates, Continents, and the Wilson Cycle	307
7.07.7.3	Plates, Climate, and Divergent Evolution of Planets	309
7.07.7.4	Plates, Super-Earths, and Habitability	309
7.07.8	**Closing and Future Directions**	311
References		311

7.07.1 Introduction

Plate tectonics is arguably one of the most successful scientific theories in physical science, given its predictive capacity regarding, for example, the distribution and magnitude of seismic and volcanic disasters (see Abbott, 2011) or seafloor ages and hydrocarbon maturation (see McKenzie, 1981). However, it also plays a crucial role in planetary science since plate tectonics is a likely key ingredient for planetary habitability. In particular the plate-tectonic mode of mantle circulation drives chemical disequilibrium in the ocean and atmosphere by constantly bringing new mantle material to the surface. In this way, plate tectonics drives the geologic carbon cycle through erosion, weathering, and volcanism (Berner, 2004; Walker et al., 1981), which imposes a negative feedback and the long-term climate stability necessary for biological evolution over billions of years. Likewise, plate tectonics may also be necessary for the origin of life by providing an energy source for chemosynthetic life at the bottom of the ocean, that is, at mid-ocean ridges (e.g., Southam and Westall, 2007) (**Chapter 10.14**). The discovery of many terrestrial planets in other solar systems over the last 15 years (e.g., Charbonneau et al., 2009) has, therefore, emphasized that the existence of plate tectonics is possibly a necessary condition for biological habitability (Foley et al., 2012; Korenaga, 2010; Landuyt and Bercovici, 2009b; O'Neill and Lenardic, 2007; Valencia and O'Connell, 2009; Valencia et al., 2007; van Heck and Tackley, 2011).

However, understanding the conditions for plate tectonics requires a predictive theory for not only how it is driven but also how it arises during the evolution of a planet and in particular how it is generated on some but not all planets. For example, the occurrence of plate tectonics on Earth but not its putative twin Venus (or any of the other terrestrial planets in our solar system) is one of the major enigmas in Earth and planetary science. Thus, a comprehensive physical theory for the origin and generation of plate tectonics is one of the greatest challenges in geophysics yet remains an elusive goal despite 30 years of research (e.g., Bercovici, 1993, 1995b; Bercovici and Ricard, 2005; Davies and Richards, 1992; Foley and Becker, 2009; Hager and O'Connell, 1979, 1981; Kaula, 1980; Landuyt and Bercovici, 2009a; Landuyt et al., 2008; Ricard and Vigny, 1989; Tackley, 1998, 2000c,d; van Heck and Tackley, 2008; Vigny et al., 1991); see previous reviews by Bercovici et al. (2000), Gurnis et al. (2000), Tackley (2000b), Bercovici (2003), Lowman (2011), and Korenaga (2013).

In the late 1930s, following the introduction of Alfred Wegener's theory of continental drift (Wegener, 1924), several driving mechanisms for the Earth's apparent surface motions were proposed. While Wegener himself favored tidal and *pole fleeing* (i.e., centrifugal) forces, Arthur Holmes and others hypothesized that thermal convection in the Earth's mantle provided the necessary force to drive continental motions (Hales, 1936; Holmes, 1931; Pekeris, 1935; see Bercovici, 2007; Hallam, 1987; Lewis, 2002; and **Chapter 7.01**). Even after the rejection of the theory of continental drift, and its resurrection 30 years later in the form of 'plate tectonics' (Hess, 1962; Le Pichon, 1968; McKenzie and Parker, 1967; Morgan, 1968; Runcorn, 1962b), mantle convection was still invoked as the engine of surface motions (e.g., Oxburgh and Turcotte, 1978; Runcorn, 1962a; Turcotte and Oxburgh, 1972).

Indeed, more than 50 years since the discovery of seafloor spreading (Hess, 1962; Vine and Matthews, 1963), there remains little doubt that the direct energy source for plate tectonics and all its attendant features (mountain building, earthquakes, volcanoes, etc.) is the release of the mantle's gravitational potential energy through convective overturn (sustained by radiogenic heating and core cooling, which partially replenishes the mantle's gravitational potential energy).

With the recognition of the importance of slab pull (e.g., Forsyth and Uyeda, 1975) and that subducting slabs are essentially cold downwellings, it is widely accepted that the plates are not so much driven by convection but are an integral part of mantle circulation, or more to the point, they *are* mantle convection (e.g., Bercovici, 2003; Davies and Richards, 1992; Schubert, 1980). However, the notion that plates arise from a convective mantle yields many complications and new questions. The most difficult of these questions concerns the property of the mantle–lithosphere system that allows a form of convective flow that is unlike most forms of thermal convection in fluids (save perhaps lava lakes; see Duffield, 1972), and even rare for terrestrial planets, except Earth, that is, a convection that looks like plate tectonics at the surface. Indeed, understanding the complex properties of the mantle's convective 'fluid' has required an extensive collaboration of disciplines including experimental petrology, mineral physics, seismology, and geodynamics.

In this chapter, we review progress on the problem of how plate tectonics itself is explained by mantle convection theory. We begin by briefly surveying some essential features of plate tectonics that geodynamic theories seek to predict or explain,

followed by aspects of basic convection in simple fluid systems that are applicable to plate tectonics (see also **Chapters 7.11** and **7.12**). We next discuss those features of plate tectonics that are reasonably well explained by basic convection theory and then examine the numerous features of plate tectonics that are poorly (or not at all) explained by simple convection. This is followed by an update on current research into the physics of plate generation and indeed how to unify the theories of plate tectonics and mantle convection. We close with a discussion of future directions regarding planetary evolution and the origin of plate tectonics itself.

7.07.2 Plate Tectonics

7.07.2.1 Present-Day Plate Motions

7.07.2.1.1 Plate kinematics

The Earth's surface motions are reasonably well described by solid-body rotations of spherical caps around an Euler pole wherein the motion of these caps is confined to the surface of a sphere (i.e., the Earth's surface). The Euler pole theory is the core premise of the working mathematical model of plate tectonics as originally prescribed by Morgan (1968) and McKenzie and Parker (1967) and continues to be used in present-day plate motion models (e.g., Argus and Gordon, 1991); (see **Chapter 6.08**). Plate-tectonic theory divides such surface motion into 12 major caps or plates, whose edges define the plate boundaries and a number of minor plate and microplates (**Figure 1**). The relative motion of these plates gives an accurate prediction of the deformation at plate boundaries, that is, whether they are convergent, divergent, or in shear. What is enigmatic about this description from a geodynamic perspective is not only that motion can be described with solid-body motion with narrow boundaries between plates but also that the Euler poles and hence directions of motion appear largely uncorrelated; for example, the motion of the Pacific and Austro-Indian Plates are essentially orthogonal to each other. These features lead to a system of plate boundaries where there is nearly as much strike-slip shear deformation as there is divergent and convergent motion (Olson and Bercovici, 1991), which is an important enigma about the plate-tectonic style of mantle convection.

Closely related to these modes of plate boundary deformation is the composition of the present-day velocity field in terms of poloidal and toroidal components (see **Section 7.07.3.7**). Poloidal flow is essentially the vertical convective circulation that expresses itself as convergent and divergent motion at the surface. Toroidal flow involves horizontal spinning or shearing motion about a vertical axis (e.g., cyclonic activity in the atmosphere), and in plate motions, this is manifested primarily as strike-slip shear. Decomposition of the present-day velocity field into poloidal and toroidal components (e.g., Forte and Peltier, 1987) shows that the poloidal and toroidal kinetic energies are comparable at most wavelengths (or spherical harmonic degree) (**Figure 2**). The largest energy for both fields occurs at spherical harmonic degree 1, for which the poloidal flow is from one hemisphere to the other and the toroidal flow involves net-lithospheric rotation relative to the mantle (Cadek and Ricard, 1992; Hager and O'Connell, 1978, 1979, 1981; O'Connell et al., 1991). Net-lithospheric rotation however depends on the choice of reference frame, the most common of which is the hot spot frame that is, in principle, fixed to the deep mantle. Although the lithosphere cannot have a net torque, net-lithospheric rotation in the hot spot reference frame is allowed and is even well predicted, given simple lateral viscosity variations associated with the contrast in oceanic and continental lithosphere (Ricard et al., 1991); in this frame of reference, the ratio of total toroidal to total poloidal kinetic energies is about 0.85, that is, the energies are roughly equal. The no-net-lithospheric-rotation frame is often chosen to remove the

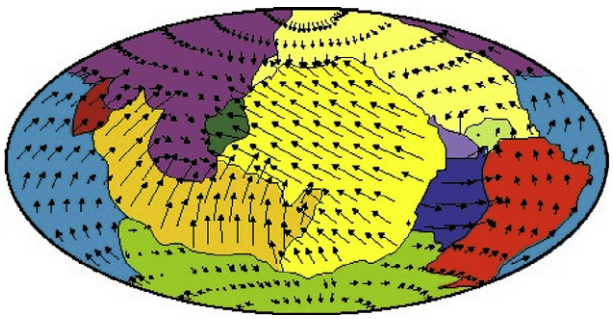


Figure 1 Configuration of present-day plate tectonics with indicated directions of motion, on an equal-area projection map. Color code for each plate is as follows: Pacific, yellow; Austro-Indian, gold; N. American, light yellow; S. American, red; African, azure; Eurasian, purple; Antarctic, chartreuse; Nazca, indigo; Cocos, violet; Caribbean, light green; Philippine, dark green; Arabian, dark red.

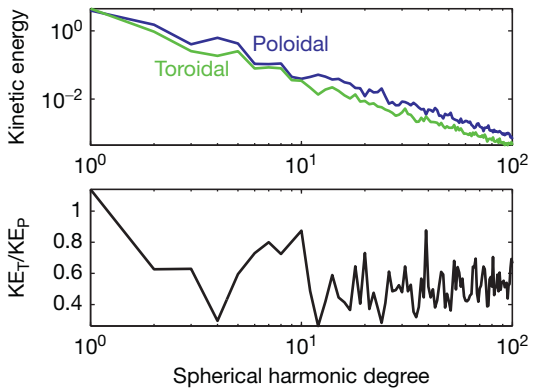


Figure 2 Power spectrum (i.e., kinetic energy per mass) of poloidal and toroidal flow fields (top frame) and the ratio of energies (bottom frame) for present-day plate motions versus the spherical harmonic degree, which is equivalently the spherical wave number (wherein larger degrees represent smaller wavelengths). Velocity can be separated into poloidal and toroidal parts according to $\mathbf{v}=\mathbf{v}_p+\mathbf{v}_t$, in which the poloidal and toroidal velocities are represented in vector spherical harmonic series $\mathbf{v}_p=\sum_{l=1}^{\infty}\sum_{m=-l}^{l}\Phi_l^m r\nabla Y_l^m$ and $\mathbf{v}_t=\sum_{l=1}^{\infty}\sum_{m=-l}^{l}\Psi_l^m \mathbf{r}\times\nabla Y_l^m$ where Φ_l^m and Ψ_l^m are spherical harmonic coefficients, r is the radial distance from center of the planet, $\mathbf{r}$ is the position vector, and Y_l^m is the normalized spherical harmonic of degree l and order m. With these expansions, the kinetic energy (per unit mass) integrated over the surface of the globe separates into poloidal and toroidal energies according to $(KE_P, KE_T)=(1/2)\sum_{l=1}^{\infty}\sum_{m=-l}^{l}l(l+1)(|\Phi_{lm}|^2, |\Psi_{lm}|^2)$. The power spectrum shows the kinetic energies per degree l, that is, the components inside the summation over l.

reference frame but is only physically plausible in a mantle with laterally homogeneous viscosity (i.e., zero torque leads to zero net rotation if the Earth is spherically symmetric); in this case, the toroidal degree-1 field is zero, and the total toroidal energy is only about 0.28 of the total poloidal kinetic energy. However, it is most instructive to examine the ratio of energies for all spherical harmonic degree, rather than lump them all together. At longest wavelength, degree-1, the energies are roughly equal. At degrees 2 and 3, typical of the Pacific basin and Ring of Fire signature, the energies are comparable with the toroidal about 60% of the poloidal. At degrees 5–10, they are similarly comparable, with the toroidal energy peaking at about 90% of the poloidal energy at degree 10, and then are comparable again at its harmonics (degrees 20 and 40); these degrees are probably indicative of average plate sizes, and the harmonics provide sharpening associated with steplike structures. Thus, while the relevant ratio of total energies is still debated because of the frame dependence, higher-order wavelengths that are representative of plate-sized structures show that the fields are comparable.

The relation of the poloidal and toroidal velocity components to plate boundaries is most easily illustrated by the rate of horizontal divergence and vertical vorticity (i.e., rate of spin and strike-slip shear), which are associated with poloidal and toroidal fields, respectively (see Figure 3). On the Earth, no plate boundary is purely poloidal or toroidal but usually some

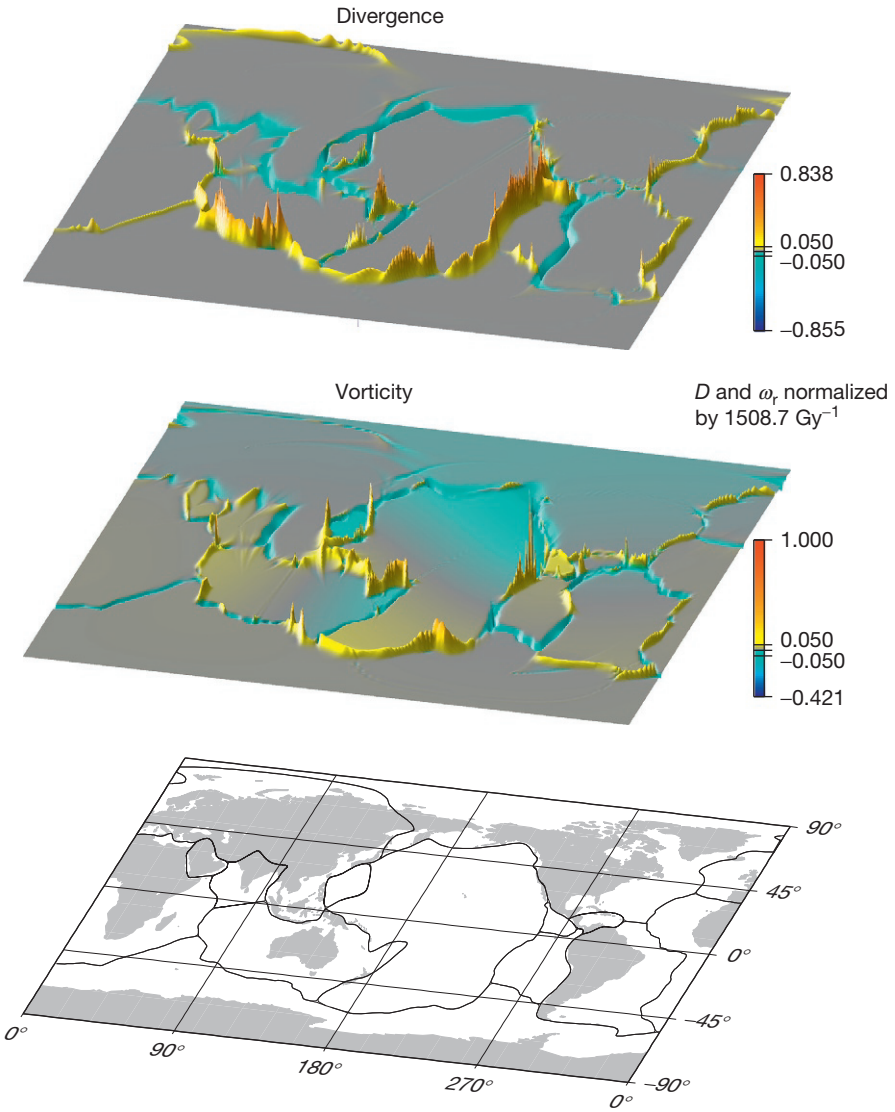


Figure 3 Approximate horizontal divergence (top) and vertical vorticity (middle) fields of the Earth (plate and continental outlines are shown for reference in the bottom frame). Horizontal divergence is $\nabla_h \cdot \mathbf{v}$ (where $\mathbf{v}$ is the surface velocity); it is a fine-scaled representation of poloidal flow at the Earth's surface and represents the so-called rate of creation (divergence) and destruction (convergence or negative divergence) of the plates at ridges and subduction zones, respectively. Vertical vorticity is $\hat{\mathbf{r}} \cdot \nabla \times \mathbf{v}$; it is a fine-scaled representation of toroidal flow and represents the rate of strike-slip shear between plates and plate spin (vorticity is essentially a measure of the local angular velocity, that is, angular velocity of a point in the fluid). The divergence and vorticity fields shown are estimated directly from a model that uses finite-width plate margins based on seismicity distributions (Dumoulin et al., 1998); for the standard plate model with infinitesimally thin margins, these fields are all predicted to be singularities.

combination. However, the associated divergence and strike-slip vorticity are both equally and very strongly localized with comparable peak values. Moreover, using the model displayed in Figure 2, the ratio of root-mean-square net vorticity to that of the divergence is about 0.7 regardless of choice of reference frame. The vorticity field displays a weak expression of plate spin associated with rotations about various Euler poles, which are visible in vorticity gradients across plate interiors; but this is very weak relative to strike-slip vorticity.

7.07.2.1.2 Plate forces

The mathematical or Euler pole theory of plate tectonics is a kinematic depiction of motion, but not causative since it does not describe the dynamics or driving forces for plate tectonics. Much of this chapter is about how the mantle convective engine is related to or even causes plate tectonics. However, a budget of plate forces can be inferred by prescribing all plate forces – for example, edge forces like slab pull, ridge push and transform resistance, and areal forces like mantle drag – that when balanced give plate velocities; this amounts to an inverse problem in which the known input data vectors are the plate boundaries, the known outputs are the plate velocities or Euler poles, and from these, one solves or inverts for the unknown forces or torques that operate on the plates to yield plate motions. The classic demonstration of this force budget was done in the seminal paper of Forsyth and Uyeda (1975). The resulting inference of plate forces allowed an assessment of how plate motions are attributed to various forces. A very clear outcome of this analysis is that plate velocities are strongly correlated with the connectivity of a plate to a slab (i.e., the percent of its perimeter taken by subduction zones), which argues rather conclusively for the dominance of slab pull as a plate driving force (Figure 4). Other features such as plate size have little correlation with plate velocity, which is either due to the lack of mantle drag or that mantle drag and distributed ridge push balance each other (Hager and O'Connell, 1981).

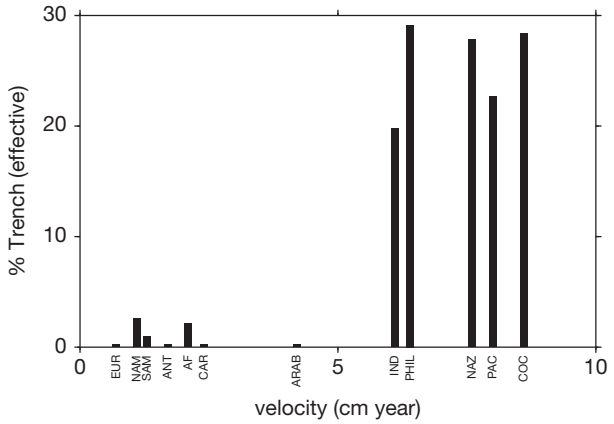


Figure 4 Percent of plate boundary that is trench or subduction zone versus plate velocity for each plate. EUR, Eurasian Plate; NAM, North American Plate; SAM, South American Plate; ANT, Antarctic Plate; AF, African Plate; CAR, Caribbean Plate; ARAB, Arabian Plate; IND, Indo-Australian Plate; PHIL, Philippine Plate; NAZ, Nazca Plate; PAC, Pacific Plate; COC, Cocos Plate. The effective trench length corrects for subduction zones whose associated slab-pull forces are opposite in direction and presumably cancel. Data from Forsyth and Uyeda (1975).

7.07.2.1.3 Plate age and size distributions

The dominance of slab pull suggests that cold downwellings drive tectonics, that is, as the plates migrate, they get colder and heavier and eventually are convectively unstable and sink (which as we will see in the succeeding text is the description of a convective thermal boundary layer). However, plates are not all old and cold at subduction zones. In particular, the seafloor age distribution implies that subduction rates are independent of plate age; that is, the age of plates at subduction zones is distributed from age nearly 0 (i.e., subducting ridges) to the oldest ages of roughly 200 My (see Becker et al., 2009; Coltice et al., 2012) (see also the discussion in the succeeding text in **Section 7.07.5.5.3**). This observation represents an important paradox in linking plate tectonics to mantle convection.

The size distribution of plates is possibly also indicative of dynamic process in the mantle and lithosphere. A cumulative size distribution of present-day plates implies two statistical populations of large plates and smaller plates that follow different power-law or fractal distribution curves (Bird, 2003; Sornette and Pisarenko, 2003). This break in the distribution also appears to be a robust feature in time, extending back to at least 200 My (Morra et al., 2013). The larger plate distribution and its evolution have been proposed to reflect the spatial and temporal scales of mantle convective processes (Morra et al., 2013), while the smaller plate population follows a standard size–frequency power or fractal law indicative of progressive fragmentation (Bird, 2003; Sornette and Pisarenko, 2003). This interpretation however is somewhat at odds with the slab-pull correlation of Forsyth and Uyeda (1975) (**Figure 4**), which suggests that fast 'active' plates driven by subduction exist in both the large and the small distributions; hence, convection does not only drive large plates, and small plates are not simply being passively fragmented between larger plates.

7.07.2.2 Origin and History of Plate Tectonics

One of the most fundamental questions regarding Earth's evolution concerns the initiation of plate tectonics itself, even though evidence for the first appearance of plate tectonics is sparse and indirect at best. Geochemical analysis of zircons implies not only the existence of liquid water (Mojzsis et al., 2001; Valley et al., 2002) more than 4 Gy ago but also melting of sediments and formation of granites, possibly related to subduction zone melting, although perhaps as likely related to repeated melting events (Harrison et al., 2005; Shirey et al., 2008), the cause for which could be tectonic style recycling. Similar evidence suggests that subduction-related crustal production existed by 3.6 Ga (Hopkins et al., 2010; Jenner et al., 2013; Polat et al., 2011). Whether such possible protosubduction was localized or global is unknown (see Korenaga, 2013, who argued that geochemical data are too undersampled to be conclusive), but geochemical and petrologic evidence implies that tectonics was not widespread until 3.0–2.7 Ga (Condie and Kroner, 2008; Shirey and Richardson, 2011); see review by van Hunen and Moyen (2012).

Aside from plate initiation, the history of plate motions has only fairly recently extended into the Proterozoic through paleomagnetic plate reconstructions (e.g., Evans, 2009; Evans and Mitchell, 2011; Evans and Pisarevsky, 2008; Torsvik et al., 1996),but has yet to reach into the Archean; thus, information about early tectonic evolution largely comes from geochemical, petrologic, and geologic history of deformation belts (e.g., Brown, 2007; Condie and Kroner, 2008; Hamilton, 1998). However, the first indications of global tectonics appear to be associated with the Wilson cycle of supercontinent aggregation and dispersal (Shirey and Richardson, 2011), which continued through the Proterozoic with supercontinents such as Nuña and Rodinia (Evans, 2009; Evans and Mitchell, 2011) and on through the Phanerozoic with continental masses such as Pangaea.

7.07.2.3 Planetary Tectonics

The lack of active plate tectonics on the other terrestrial planets of our solar system is an important constraint for inferring the conditions under which plate tectonics does or does not exist. Although Venus is very similar in size to Earth, and should thus have similar internal dynamics, it does not have recognizable plate tectonics. However, Venus' dearth of craters implies it was extensively resurfaced about 500 My ago, by either volcanic activity or lithospheric foundering (Turcotte, 1993; Turcotte et al., 1999). Moreover, the arcuate and topographic signatures of large Venus coronae are indicative of possible subduction, although other types of tectonic boundaries are not evident (Schubert and Sandwell, 1995). Although Mars is considerably smaller than Earth and thus more likely to have a colder and less active interior, it might have had tectonic activity in its early history; in particular, magnetic features and geologic reconstructions indicated the possibility of ancient spreading centers and strike-slip boundaries (Connerney et al., 1999, 2005; Yin, 2012). The lack of plate tectonics on Mars and Venus is typically attributed to the lack of liquid water, since Venus's water has been lost through a runaway greenhouse and Mars' water only likely existed in liquid form in its early history (Squyres et al., 2004), though perhaps coincident with surface tectonics. The mechanism by which liquid water provides a condition for plate tectonics remains enigmatic and debated since sustaining pore pressures, which reduce friction, to depths of 100 km is highly problematic, and the Earth's lithosphere might be as dry as Venus' due to dehydration melting at ridges (Hirth and Kohlstedt, 1996). Moreover, since plate tectonics is deemed necessary to stabilize Earth's temperate climate by a negative weathering feedback (Berner, 2004; Walker et al., 1981), whether liquid water or protoplate tectonics appeared first after magma ocean solidification remains an open question (e.g., Driscoll and Bercovici, 2013; Hamano et al., 2013; Lebrun et al., 2013).

7.07.3 Basic Convection

Basic thermal convection in fluids is perhaps the most fundamental paradigm of self-organization in nonlinear systems (see Nicolis, 1995). Such self-organization is characterized by the forcing of simple homogeneous yet perturbable or mobile systems (i.e., whose particles can move) far from equilibrium (e.g., by heating or imposition of chemical disequilibrium); in many cases, the systems can develop complex patterns and oscillatory or chaotic temporal behavior. In basic thermal convection, for example, a layer of fluid is heated uniformly and subsequently develops organized polygonal patterns and cells of cold and hot thermals. Indeed, the formation of tectonic plates is invariably a form of convective self-organization.

Although the phenomenon of convection was first recognized by Benjamin Thompson (Count Rumford) (1797), the first systematic experimental study of basic convection was carried out by Henri Bénard on layers of spermaceti (whale fat) and paraffin (Bénard, 1900, 1901), which yielded striking images of honeycombed patterns and vertical cellular structure (see **Chapter 7.03**). After John William Strutt (Lord Rayleigh) (1916) failed to theoretically explain the experiments with hydrodynamic stability theory, it was eventually deduced that Bénard's experiments were influenced by surface tension because of their open surface (Block, 1956; Pearson, 1958). When the thermal convection experiments were carried out with more controlled conditions, they were found to be well predicted by Rayleigh's theory (see Chandrasekhar, 1961, Chapter 2, Section 18). Nevertheless, convection in a layer of fluid heated along its base is still called Rayleigh–Bénard (and often just Bénard) convection (See **Chapter 7.01**).

7.07.3.1 Convective Instability and the Rayleigh Number

Rayleigh's stability analysis of convection predicted the conditions necessary for the onset of convection, as well as the expected size of convection cells (relative to the layer thickness). The mathematics of his analysis has been summarized well elsewhere (Chandrasekhar, 1961), but in essence, the theory determines the infinitesimal thermal perturbations most likely to trigger convective overturn in a fluid layer that is gravitationally unstable (i.e., hotter, and thus less dense, on the bottom than the top). This most destabilizing perturbation is called the least-stable or most-unstable mode; its prediction and analysis are fundamental to the study of pattern selection in convection, as well as other nonlinear systems (see **Chapters 7.02** and **7.04**).

Depending on the nature of the binding top and bottom surfaces of the layer (i.e., whether they are freely mobile or adjacent to immobile surfaces), the most-unstable mode leads to convection cells that have a width approximately equal to (though as much as 50% larger than) the depth of the fluid layer. This result is quite relevant to plate tectonics since it is the first basic step in predicting the size of convection cells and thus the size of plates.

The perturbation that leads to this form of convection is deemed most unstable because, of all the possible perturbations, it will induce overturn with the smallest forcing. The forcing (i.e., deviation from conductive equilibrium) is driven by imposed heating, such that the resultant thermal buoyancy (proportional to density contrast times gravity) of a hot fluid parcel rising from the bottom surface through colder surroundings acts to drive convective overturn. However, such motion is also resisted or damped in two unique ways: viscous drag acts to slow down this parcel, and thermal diffusion acts to erase its

hot anomaly (i.e., to lose heat to its colder surroundings). Thus, while the fluid layer might be gravitationally unstable, hot parcels rising might move too slowly against viscous drag before being erased by thermal diffusion. Similar arguments can be made for cold material sinking from the top surface through warmer surroundings. The competition between forcing by thermal buoyancy and damping by viscosity and thermal diffusion is characterized in dimensionless ratio called the Rayleigh number:

$$Ra = \frac{\rho g \alpha \Delta T d^3}{\mu \kappa} \quad [1]$$

where ρ is fluid density, g is gravitational acceleration, α is thermal expansivity (units of K^{-1}), ΔT is the difference in temperature between the bottom and the top surfaces, d is the layer thickness, μ is fluid viscosity (units of Pa s), and κ is fluid thermal diffusivity (units of $m^2 s^{-1}$).

Even though $\Delta T > 0$ (i.e., heating is from below and causes gravitational instability), Ra still must exceed a critical value Ra_c for convection to occur. For $Ra < Ra_c$ the layer is stable and transports heat by conduction; for $Ra > Ra_c$, the layer will be convectively unstable and transport heat more rapidly via convection. Although Ra_c varies depending on the mechanical nature of the horizontal boundaries (whether rigid or a free surface), it is typically of order 1000 (i.e., it varies between $Ra_c \approx 700$ for freely slipping – called *free-slip* – top and bottom boundaries and $Ra \approx 2000$ for the rigid or *no-slip* top and bottom boundaries).

This value can be roughly understood by considering the ascent of a hot (or descent of a cold) parcel of size a and temperature anomaly ΔT against a viscous medium. Dimensional analysis shows that the parcel's typical ascent velocity is $\rho g \alpha \Delta T a^2 / \mu$ (i.e., this combination of constants has units of meter per second). However, the rate that heat diffuses out of the parcel is κ / a (i.e., smaller parcels lose heat faster). The critical state occurs when these two rates are equal; that is, if the buoyant ascent rate just exceeds the diffusion rate, the parcel should rise without being erased, but if the ascent rate is less than the diffusion rate, it will barely rise before being lost. Therefore, the critical state occurs if $\rho g \alpha T a^3 / (\mu \kappa) \approx 1$. Scaling purely by orders of magnitude, a small parcel of fluid can be assumed to be of order ten times smaller than the entire layer; thus, assuming $a \approx d/10$ leads to a critical condition for onset of convection of $\rho g \alpha \Delta T d^3 / (\mu \kappa) \approx 1000$.

For the Earth's mantle, the typical average properties from which the Rayleigh number is constructed are $\rho \approx 4000$ kg m^{-3}, $g = 10$ m s^{-2}, $\alpha = 3 \times 10^{-5}$ K^{-1}, $\Delta T \approx 3000$ K, $d = 2900$ km, $\mu = 10^{22}$ Pa s (dominated by the lower mantle), and $\kappa = 10^{-6}$ m^2 s^{-1} (see Schubert et al., 2001) (see **Chapter 7.02** and Volume 2). Taken together, these lead to a Rayleigh number of $\sim 10^7$, which is well beyond supercritical; although the mantle viscosity is extremely high, the mantle is also very hot and very large and hence convects vigorously. (See **Chapters 7.02** and **7.04** for more thorough discussion.)

7.07.3.2 Vertical Structure of Simple Convection

7.07.3.2.1 Symmetry

Once convection initiates and becomes fully developed, it is called *finite amplitude* convection, as opposed to the infinitesimal convection discussed earlier. The simplest form of such convection involves two-dimensional flow of constant-viscosity fluid in a plane layer that is heated on the bottom and cooled on the top (but whose upper and lower boundaries are otherwise mechanically identical, i.e., both free-slip or both no-slip). Such convection is a highly symmetrical system. This symmetry is prescribed by the fact that when convection reaches its final state (i.e., a steady state or time-averaged steady state), the top boundary is cooling the fluid by the same amount that the bottom boundary is heating the fluid; otherwise, the layer would heat up or cool down and is thus not in its final state. Thus, cold currents sinking from the top boundary will generally be equal and opposite in temperature and velocity to the hot currents emanating from the bottom boundary (**Figure 5**). The vertical cross section of such two-dimensional (2-D) steady convection therefore shows sequences of mirror image pairs of counterrotating convection cells; each cell's vertical wall is either a hot upwelling or a cold downwelling current, and the upwellings and downwellings are themselves symmetrical mirror images of each other through a 180° rotation (**Figure 5**). The breaking of this symmetry by various effects is vital to understanding the plate-tectonic style of mantle convection.

7.07.3.2.2 Thermal boundary layers

When basally heated convection is vigorous (i.e., has moderate to high Rayleigh number Ra), it tends to stir the fluid until it is nearly homogeneous, that is, most of the fluid is at or near the average temperature of the top and bottom boundaries (this interior temperature is close to isothermal unless the fluid is significantly compressible over the depth of the layer, in which

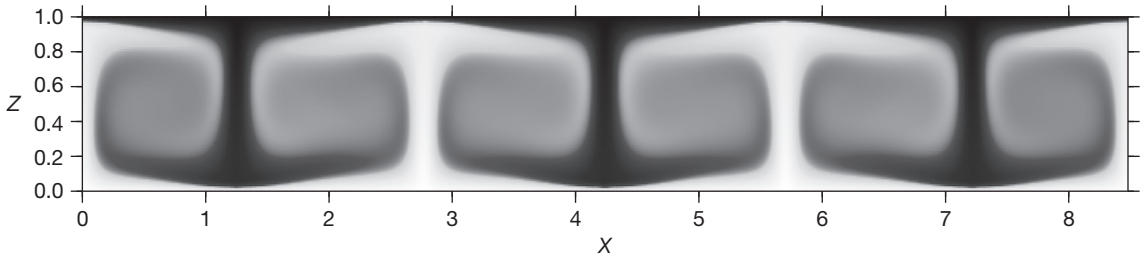


Figure 5 A vertical cross section of the temperature field of a numerical simulation of two-dimensional plane-layer, basally heated, isoviscous (Bénard), convection, with free-slip top and bottom boundaries and $Ra = 10^5$. Black represents cold fluid; light gray is hot fluid. The temperature field shows symmetrical convection cells, upwellings, downwellings, and thermal boundary layers thickening in the direction of motion (at the top and bottom of the layer, in between the upwellings and the downwellings).

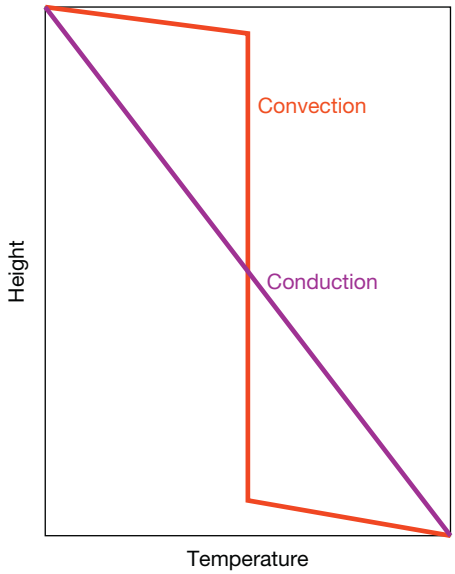


Figure 6 Sketch of temperature profiles (i.e., horizontally averaged temperature vs. depth) for a simple isoviscous basally heated plane layer. Convective mixing homogenizes the conductive mean temperature into a nearly isothermal state (if the fluid is incompressible) with thermal boundary layers connecting it to the cold surface and hot base.

case the interior temperature follows an adiabat; see Chapter 7.02), thereby relieving the gravitational instability for the much of the layer (Figure 6). However, directly adjacent to the top and bottom boundaries, the fluid temperature must make a transition from the largely homogeneous interior value to the colder (at the top) or hotter (at the bottom) boundary value (again see Figure 6). These transition regions are the *thermal boundary layers*. In effect, convection confines the gravitationally unstable parts of the fluid to these relatively thin boundary layers, keeping most of the rest of the fluid gravitationally stable. The Earth's lithosphere and tectonic plates are essentially the horizontal thermal boundary layer along the top surface of the Earth's convecting mantle; thus, we will discuss thermal boundary layers in some detail and refer to them repeatedly.

Because of their large vertical temperature gradients, the thermal boundary layers control, via thermal conduction, the influx (through the bottom) and efflux (through the top) of heat. If the fluid is stirred by convection with increasing vigor, then more of the fluid will become homogeneous, and thus, the boundary layers will become thinner; although this leads to yet more gravitationally stable material, it also causes sharper thermal gradients in the boundary layers and thus even larger heat fluxes in to and/or out of the layer. Therefore, the more vigorous the convective stirring, the greater the heat flow through the layer. (For basally heated convection in which the top and bottom boundaries are isothermal, the heat flow through the layer is, in theory, unbounded; that is, convection can be so vigorous that the fluid would appear isothermal with two thermal boundaries reduced to simple temperature discontinuities. The same is not true for convection with only internal heating (see succeeding text) since the heat flow must balance the net rate of heat production in the layer regardless of fluid properties.)

As mentioned, the thermal boundary layers are where the gravitationally unstable material is confined, and thus, fluid in these layers must eventually either sink (if in the cold top boundary layer) or rise (if in the hot bottom boundary layer), thereby feeding the vertical convective currents, that is, downwellings and upwellings, respectively. The feeding of vertical currents induces motion of the boundary layers toward convergent zones (e.g., a downwelling for the top boundary layer) and away from divergent zones (e.g., over an upwelling). Such flow naturally causes the boundary layers to thicken in the direction of motion (see Figure 5). For example, in the top cold thermal boundary layer, fluid from hotter depths newly arrived at the divergent zone heats and thus thins the boundary layer; but as the fluid moves toward the convergent zone, it cools against the surface, and the boundary layer gradually thickens before eventually growing heavy enough to sink into the downwelling. We will revisit boundary layer thickening again when discussing subduction zones and seafloor topography.

7.07.3.2.3 The size of convection cells

As discussed in Section 7.07.3.1, stability theory predicts that at the onset of overturn, convection cells are essentially as wide as the layer is thick. What determines the lateral extent of a single convection cell in finite amplitude convection is the length of the horizontal boundary layer currents. For example, as material in the top boundary layer flows laterally, it loses heat and buoyancy (being adjacent to the cold surface) and thus travels only a certain distance before it becomes so cold and negatively buoyant that it sinks and feeds the cold downwelling. The horizontal distance this boundary layer material travels before sinking determines the size of convection cells. For simple fluids, convection cell width is predicted by boundary layer theory to be the same as the layer height, that is, with an aspect ratio (width to height) of unity (Turcotte and Oxburgh, 1967; Turcotte and Schubert, 2002), as verified by experiments (e.g., see Figure 7 similar boundary layer arguments can be used to understand the structure of tectonic plates, although the aspect ratios of the major plates are much greater than unity, which reflects rheological complexities (see Section 7.07.3.5).

7.07.3.2.4 Thermal boundary layer forces

What force makes the thermal boundary layer currents flow laterally? While it seems obvious that boundary layer fluid feeding a vertical current must flow horizontally (such as fluid going toward a drain), the forces behind these currents are of significance in regard to plate forces. First, buoyancy does not drive the boundary layer currents directly; buoyancy only acts vertically, while boundary layer currents move horizontally (buoyancy or gravity does eventually deflect these currents back into the convecting layer, but it cannot drive their lateral flow). Horizontal pressure gradients are the primary driving force for thermal boundary layers. For example, when hot upwelling fluid impinges on the top boundary, it induces a high-pressure region below the surface. As the top cooling thermal boundary layer material flows away from that region to its own downwelling, it gets heavier and acts to pull away from the surface; this induces a suction effect and thus, because of the boundary layer's growing weight, increasingly

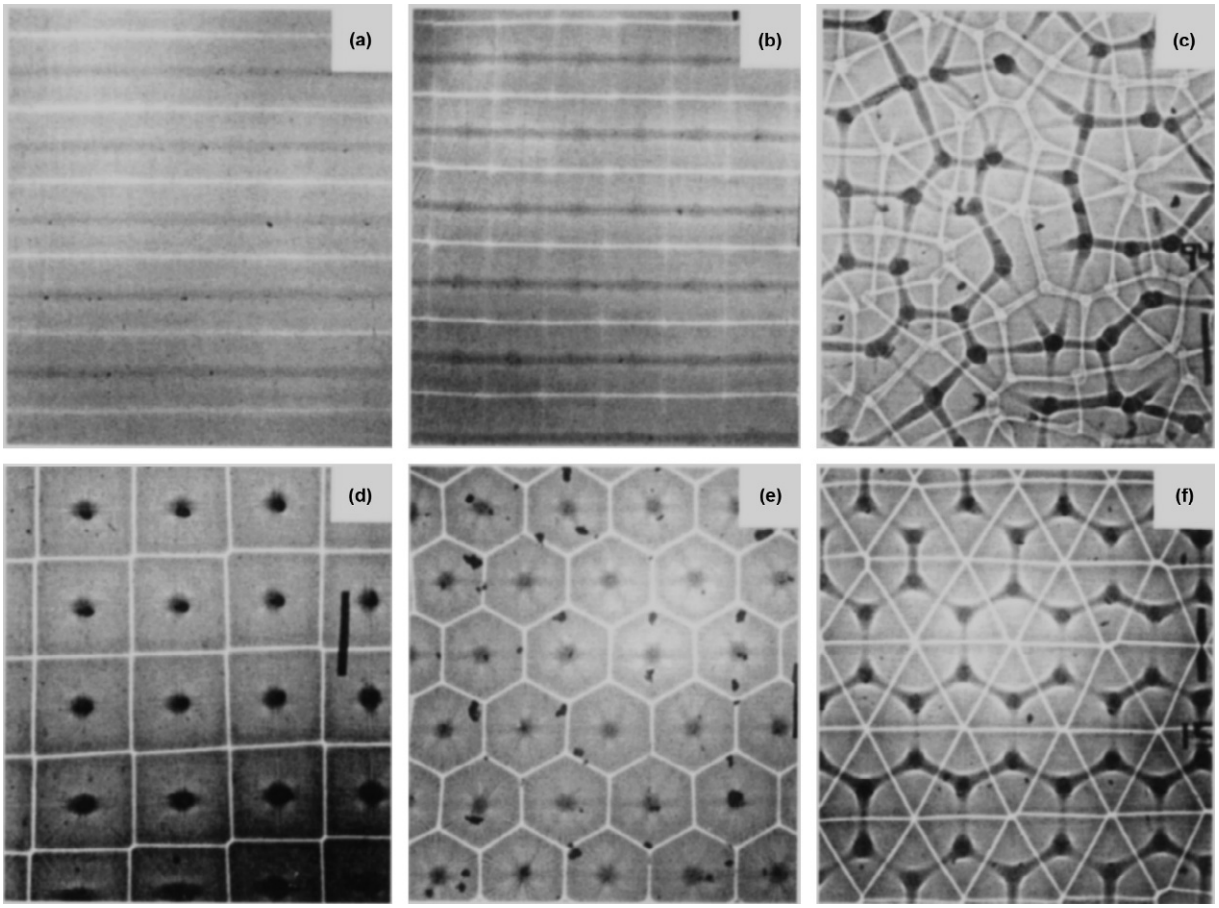


Figure 7 Laboratory experiments on convection in viscous liquids in a plane layer, bounded between two horizontal rigid glass plates and heated from below. All images shown employ the shadowgraph technique (see text for discussion), and thus, dark regions are relatively hot, while light regions are colder. The roll pattern (a) is also called 2-D convection; this pattern becomes unstable to the bimodal pattern (b) at moderate Rayleigh numbers, which eventually gives way to the spoke pattern (c) at yet higher Rayleigh numbers (around $Ra = 10^5$). The roll, bimodal, and spoke patterns are typical for convection in isoviscous fluids, although they also occur in fluids with temperature-dependent viscosity ((a)–(c) are in fact for weakly temperature-dependent viscosity fluids). More unique to convection in temperature-dependent viscosity fluids are squares (d), hexagons (e), and even triangles (f). Reproduced from White D (1988) The planforms and onset of convection with temperature-dependent viscosity. *Journal of Fluid Mechanics* 191: 247–286.

lower pressures in the direction of motion, eventually culminating in a concentrated low-pressure zone where the downwelling separates from the surface. Thus, the horizontal boundary layer current flows from the induced high pressure over the upwelling to the low pressure over the downwelling, that is, it flows down the pressure gradient. Invariably plate driving forces are related to these pressure gradients. Ridge push is the gradual pressure gradient going from the pressure high at a ridge outward; slab pull is effectively due to a concentrated pressure low caused by slabs pulling away from the surface. We discuss plate forces in greater detail later (**Section 7.07.4.1**).

7.07.3.3 Patterns of Convection

The 2-D convection described earlier is a special type of convective flow also known as convection with a roll planform; that is, the 2-D convection cells are infinitely long counter-rotating cylinders or rolls when extended into three dimensions. This pattern of convection is typically unstable at moderately high Rayleigh numbers (i.e., somewhat greater than the critical *Ra*) at which point convection becomes 3-D (Busse and Whitehead, 1971); that is, roll-like convection cells break down into more complicated geometric shapes. The study of convective planforms and pattern selection is a very rich and fundamental field in itself, although a full quantitative discussion is beyond the scope of this chapter; the interested reader should refer to Busse (1978) (and **Chapter 7.03**). However, we will briefly survey convective patterns in the context of plate tectonics and plate generation.

Much of the work on convective planforms was motivated by experimental studies of convection in thin layers (e.g., see Busse, 1978; Busse and Whitehead, 1971; Whitehead, 1976; Whitehead and Parsons, 1978) (see **Chapter 7.03**). Convection patterns displayed with experimental (and computational) techniques are quite varied. Purely basally heated, plane-layer isoviscous convection is a naturally symmetrical system; thus, even when convection is 3-D, the upwelling and

downwelling currents retain some basic symmetry; for example, in *bimodal* convection, the upwellings assume the geometry of two adjacent sides of a square, while the downwellings form the other two sides of the square (Figure 7). However, the convection planform can also become highly irregular as with the *spoke* pattern that shows several nearly linear upwellings joined at a common vertex (and likewise with downwellings) (Figure 7); even in this irregular planform, the upwellings and downwellings are symmetrical in that, apart from the sign of their thermal anomalies, their general structure is indistinguishable, that is, they have the same irregular spoke shapes.

Other patterns of convection are also possible but are mostly seen with nonconstant-viscosity fluids. Even in this case, regular polyhedral patterns like squares and hexagons are common (as are irregular forms like the spoke pattern). One of the basic rules of the formation of regular planforms is that they obey a 'tiling pattern'; that is, if the fluid layer is to be filled with identical convection cells, then the cells must all fit together precisely (else a misfit would constitute a cell with an aberrant shape). There are only a few geometries that will allow identical cells to 'tile' together, that is, infinitely long rolls, rectangles, squares, triangles, and hexagons (see Figure 7). Even regular cells with hexagonal or square planforms are characteristic of weakly temperature-dependent viscosity and involve asymmetry between upwellings and downwellings, as we will discuss in the succeeding text (Section 7.07.3.5). Indeed, the symmetry breaking of basally heated, isoviscous convection by various effects, most notably internal heating and variable viscosity, is fundamental to understanding the nature of mantle convection and its relation to plate tectonics.

7.07.3.4 Influence of Internal Heating

As discussed earlier, purely basally heated convection in plane layers leads to upwellings and downwellings of equal intensity (equal and opposite thermal anomalies and velocities). However, one of the defining characteristics of convection in the Earth's mantle is that it is powered by the release of primordial heat, left over from planetary accretion, and radiogenic heating from the decay of uranium, thorium, and potassium distributed throughout the mantle. Thus, in addition to being heated along the core–mantle boundary by the hotter molten iron outer core, the mantle retains fossil heat while also being heated throughout its interior. Both primordial heat and radiogenic heating are referred to as internal heating – in contrast to boundary heating – and they effectively break the symmetry between upwellings and downwellings (Figure 8).

The simplest form of convection driven by internal heating is one in which the heat sources are distributed uniformly, the top boundary is kept cold and isothermal, and the bottom boundary is thermally insulated, that is, no heat passes through it; this is called purely internally heated convection. In this case, the bottom boundary cannot develop a thermal boundary layer, as this would carry heat-conducting thermal gradients, which are disallowed by the insulating condition of the boundary. Since there is no hot bottom thermal boundary layer to provide positive buoyancy, then there are no focused active upwellings emanating from the base of the layer. However, the top boundary is cold and thus develops a thermal boundary layer whose temperature gradients are responsible for conducting out all of the internally generated heat. Material in this boundary layer is cold and heavy and is pulled horizontally (again, via pressure gradients) toward a convergent zone where it eventually sinks to form cold downwellings, which cool the interior of the fluid layer. Thus, in purely internally heated convection, there are only concentrated downwelling currents; to compensate for the resulting downward mass flux, upwelling motion occurs, but it tends to be a broad background of diffuse flow, rising passively in response to the downward injection of cold material. Therefore, since the downwellings descend under their own negative buoyancy, they are typically called *active* currents, while the background upwelling is deemed *passive*.

Convection with both internal heating and basal heating (i.e., the bottom boundary is hot and isothermal – instead of insulating – and thus permits the passage of heat) is more complex than with either pure basal heating or pure internal heating. However, the nature of the resulting convection can be understood if one realizes that the bottom thermal boundary layer must conduct in the heat injected through the bottom, while the top thermal boundary layer must conduct out both the heat injected through the bottom and the heat generated internally (Figure 8). To accommodate this extra heat flux, the top thermal boundary layer develops a larger temperature drop

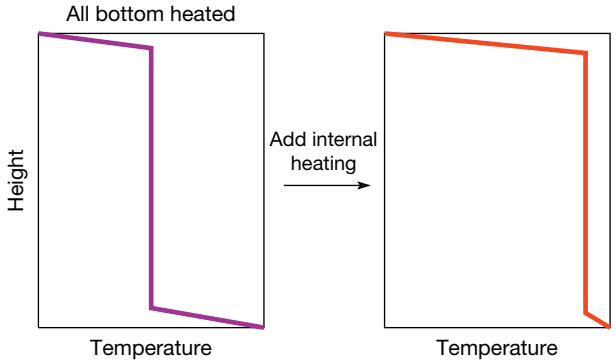


Figure 8 Sketch of temperature profiles, as in Figure 6, for an isoviscous plane layer, with and without internal heating. With no internal heating, the interior mean temperature is the average of the top and bottom temperatures; the effect of adding internal heating is to increase the interior mean temperature and thus change the relative size and temperature drop across the top and bottom thermal boundary layers.

(to sustain a larger thermal gradient) than does the bottom boundary layer. In this way, the top boundary layer has a larger thermal anomaly than the bottom one, leading to larger, more numerous, and/or more intense cold downwellings than hot upwellings. Invariably, internal heating breaks the symmetry between upwellings and downwellings by leading to a preponderance of downwellings driving convective flow. In the case of Earth, the net surface heat flux is traditionally thought to be as much as 70–80% due to internal heating (with comparable parts radiogenic and primordial heat) and 20% basally heated (Turcotte and Schubert, 2002), although recent estimates of large thermal conductivity of liquid iron in the outer core imply a larger core heat flux and a mantle that is more than 25% basally heated. The mantle is, in any case, predominantly heated internally, with a smaller amount of basal heating. Thus, one can expect a top thermal boundary layer with a large temperature drop across it feeding downwellings that dominate the overall convective circulation; active upwellings from the heated bottom boundary provide only a small amount of the net outward heat flux and circulatory work. Indeed, this picture of convective circulation is borne out by both laboratory and numerical experiments (Figure 9).

Although we have yet to discuss many of the complexities leading to the plate-tectonic style of mantle convection, this simple picture of internally heated convection is in keeping with the idea that the large-scale circulation is driven by downwellings (slabs) fed by an intense thermal boundary layer (the lithosphere and plates), while active upwellings (mantle plumes) are relatively weak and/or few in number (Bercovici et al., 1989a; Bunge et al., 1997; Davies and Richards, 1992).

7.07.3.5 Influence of Temperature-Dependent Viscosity

Mantle material is known to have a highly temperature-dependent viscosity for subsolidus, or solid-state, flow.

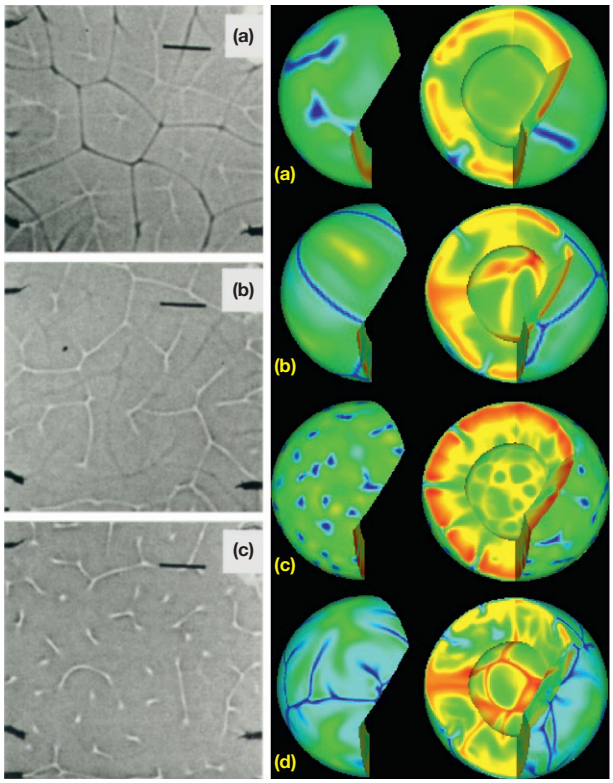


Figure 9 Internally heated convection in examples of laboratory (left) and numerical (right) experiments. Laboratory experiments employ the shadowgraph technique (see Figure 7 and text for discussion), and apparent internal heating is accomplished by initiating basally heated convection and then steadily decreasing the temperature of the top and bottom surfaces; this causes the average fluid temperature to be greater than the average temperature of the two boundaries. The fluid thus loses its net heat and the rate of this bulk cooling is a proxy for internal heating. The frames show the convective pattern when the layer is all basally heated (a), and thus, upwellings (dark zones) and downwellings (light zones) are comparable; when bulk cooling (internal heating) is stronger than basal heating, the downwellings (light) are dominant (b); when bulk cooling is much stronger than basal heating, the upwellings are not distinct enough to register in the shadowgraph (c) (adapted from Weinstein S and Olson P (1990) Planforms in thermal convection with internal heat sources at large Rayleigh and Prandtl numbers. *Geophysical Research Letters* 17: 239–242). Numerical experiments are for purely internally heated convection in a spherical-shell model of the mantle with free-slip top and bottom surfaces, at moderate and high *Ra*, with and without a viscosity jump at 660 km depth. (a) Isoviscous mantle at $Ra = 1.6 \times 10^5$. (b) Same as (a) but with the viscosity of the upper mantle (above 660 km) reduced by a factor of 30. (c) Same as (a) except the entire mantle viscosity is reduced by 30 (thus, *Ra* is larger by 30). (d) Same as (b) except the viscosity of both layers is reduced by 10. Note that even while an increase in *Ra* causes more columnar downwellings (from (a) to (c)), a viscosity jump tends to reinforce or restore the tendency for sheetlike downwellings (d). Adapted from Bunge H, Richards M, and Baumgardner J (1996) Effect of depth-dependent viscosity on the planform of mantle-convection. *Nature* 379: 436–438.

Whether such subsolidus flow occurs by diffusion creep (deformation through the diffusion of atoms away from compressive stresses toward tensile stresses) or dislocation power-law creep (dislocations in the crystal lattice propagate to relieve compression and tension), the mobility of the atoms under applied stresses depends strongly on thermal activation; that is, the atom's thermal kinetic energy determines the probability that it will jump out of a lattice site (Evans and Kohlstedt, 1995; Karato, 2008; Ranalli, 1995; Weertman and Weertman, 1975). The viscosity law for silicates therefore contains a quantum-mechanical probability distribution in the form of the Arrhenius factor $e^{H_a/RT}$ where H_a is the activation enthalpy (basically the height of the energy potential well of the lattice site out of which the mobilized atom must jump), R is the gas constant, T is temperature, and thus RT represents the average kinetic energy of the atoms in the lattices sites. Because of this factor, a few hundred degree change in temperature can cause many orders of magnitude change in viscosity. Moreover, with the inverse dependence on T in the Arrhenius exponent, viscosity is highly sensitive to temperature fluctuations at lower temperatures (i.e., the viscosity vs. temperature curve is steepest at low T). Thus, in the coldest region of the mantle, that is, the lithosphere, viscosity undergoes drastic changes: mantle viscosity goes from 10^{21} Pa s in the lower part of the upper mantle, to as low as 10^{18} Pa s in the asthenosphere (e.g., King, 1995), to an effective value of 10^{25} Pa s (e.g., Beaumont, 1976; Watts et al., 1982) or potentially much higher in the lithosphere. Thus, the viscosity may change by at least seven orders of magnitude in the top 200 hundred kilometers of the mantle. In the end, the effect of temperature-dependent viscosity on mantle convection is to make the top colder thermal boundary – that is, the lithosphere – much stronger than the rest of the mantle. This phenomenon helps make thermal convection in the mantle platelike at the surface in some respects, but it can also make convection *less* platelike in other respects.

A strong temperature-dependent viscosity can break the symmetry between upwellings and downwellings in much the same way as internal heating. This occurs because the top cold thermal boundary layer is mechanically much stronger and less mobile than the hotter bottom thermal boundary layer. The less mobile upper boundary induces a heat plug that forces the fluid interior to warm up; this in turn causes the temperature contrast between the fluid interior and the surface to increase and the contrast between the fluid interior and underlying medium (the core in the Earth's case) to decrease. This effect leads to a larger temperature jump across the top boundary layer (which partially reduces the boundary layer's stiffness by increasing its average temperature), and a smaller one across the bottom boundary layer (see **Figure 10**). Thus, the temperature-dependent viscosity can lead to an asymmetry in the thermal anomalies of the top and bottom boundary layers much as we see in the Earth.

The asymmetry of the boundary layers and related warming of the convective layer's interior is also related to the classical concept of 'self-regulation' in terrestrial planets (Tozer, 1972). In particular, if the mantle viscosity is too high for convection to be vigorous enough to remove the internally generated heat, then the mantle will simply heat up until the viscosity is reduced sufficiently to increase convective vigor. This effect plays an important role in the Earth's thermal evolution (Davies, 1980; Schubert et al., 1980), although it might be less profound if the radiogenic internal heating is less than originally thought and closer to that predicted by chondritic abundances (Korenaga, 2008).

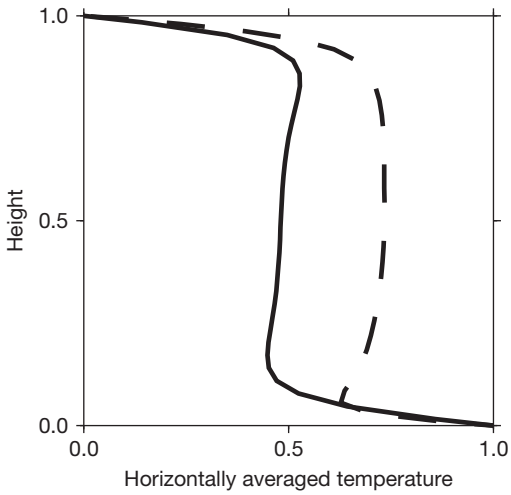


Figure 10 Temperature profiles, as in **Figure 6**, for a basally heated plane layer of fluid undergoing thermal convection when its viscosity is constant (solid curve) and temperature-dependent (dashed). The fluid with temperature-dependent viscosity develops a stiffer upper thermal boundary layer that acts as a heat plug (i.e., it reduces convection's ability to eliminate heat), causing most of the rest of the fluid to heat up to a larger average temperature. Adapted from Tackley P (1996) Effects of strongly variable viscosity on three-dimensional compressible convection in planetary mantles. *Journal of Geophysical Research* 101: 3311–3332.

Temperature-dependent viscosity can also cause a significant change in the lateral extent of convection cells. Because material in the top thermal boundary must cool a great deal and thus for a long time to become negatively buoyant enough to sink against its cold, stiff surroundings, it must travel horizontally a long distance while waiting to cool sufficiently, assuming it travels at a reasonable convective velocity; this can cause the upper thermal boundary layer and thus its convection cell to have excessively large lateral extents relative to the layer depth (i.e., large *aspect ratios*). This effect has been verified in laboratory and numerical experiments (**Figure 11**) (Giannandrea and Christensen, 1993; Ratcliff et al., 1997; Tackley, 1996a; Weinstein and Christensen, 1991). As will be discussed later (**Section 7.07.4.6**), large aspect ratio convection cells are considered to occur in the Earth, especially if one assumes that the Pacific Plate and its subduction zones reflect the dominant convection cell in the mantle (i.e., a width of ~15 000 km for a maximum depth of ~3000 km). Thus, temperature-dependent viscosity can be used to explain the large aspect ratio of convection cells of mantle convection but with some significant qualifications.

With very strong temperature-dependent viscosity, the top thermal boundary layer can also become completely immobile and the large aspect ratio effect vanishes. The immobile boundary layer happens simply because it is so strong that it cannot move. As a result, the top boundary layer effectively imposes a rigid lid on the rest of the underlying fluid, which then

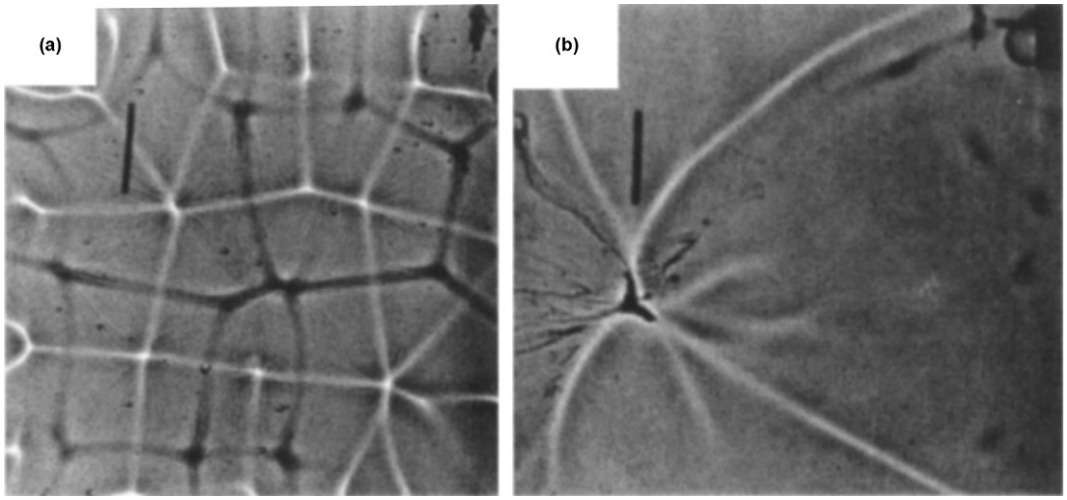


Figure 11 Lab experiment for convection with temperature-dependent viscosity fluid and mobile top thermal boundary layer. The mobility of the top boundary layer is facilitated by a free-slip upper boundary, accomplished by inserting a layer of low-viscosity silicon oil between the working fluid (corn syrup) and the top cold glass plate. Without the oil, the rigidity of the top glass plate tends to help immobilize the cold and stiff top boundary layer. (a) For comparison, the experiment when the top boundary is rigid, yielding a predominantly spoke-like pattern. (b) The planform when the oil is used to make the free-slip top boundary, causing a dramatic increase in the convection cell size to what is deemed the *spider* planform. Adapted from Weinstein S and Christensen U (1991) Convection planforms in a fluid with a temperature-dependent viscosity beneath a stress-free upper boundary. *Geophysical Research Letters* 18: 2035–2038.

convects much as if it were nearly in isoviscous convection with a no-slip top boundary condition. Convection then has cells that are again about as wide as they are deep (i.e., with nearly unit aspect ratio). Moreover, in this form of convection, the planform can assume various simple geometries such as squares and hexagons (**Figure 7**); however, the upwellings and downwellings are not symmetrical in that the downwellings form the edges of the hexagons (or squares) and the hot, low-viscosity upwellings tend to rise as pipes or plumes through the center of the hexagons (or squares) (see Busse, 1978), similar in many respects to plumes in the Earth's mantle. Regardless of the pattern of convection, the immobilization of the top thermal boundary layer leads to convection that is unlike the Earth: unit aspect ratio cells and an immobile lithosphere.

Overall, three different styles or regimes of convection with temperature-dependent viscosity were first noted by Christensen (1984b) and elucidated and summarized by Solomatov (1995) with a basic scaling analysis (**Figure 12**). These regimes are as follows: (1) For weakly temperature-dependent viscosity, convection tends to appear as if it were nearly isoviscous (unit aspect ratio cells and a mobile top boundary); (2) with moderately temperature-dependent viscosity, convection develops a sluggish cold top boundary layer that is mobile but with large horizontal dimensions, that is, large aspect ratio cells; and (3) with strong temperature-dependent viscosity, the top boundary layer becomes rigid and immobile and convection assumes much of the appearance (apart from detailed patterns) of isoviscous convection beneath a rigid lid. The size of these regimes depends on the Rayleigh number since vigorous convection is more capable of mobilizing a cold top boundary layer than is gentle convection. These regimes have been deemed the nearly isoviscous or low-viscosity contrast regime, the sluggish convection regime, and the stagnant-lid regime. Although the sluggish regime is characterized by large aspect ratio cells suggestive of the Earth's plates, it is most likely that the Earth's Rayleigh number and viscosity variability would put the Earth's mantle in the stagnant-lid regime, that is, if the mantle were a fluid with only temperature-dependent viscosity. Indeed, both Venus and Mars, lacking unequivocal signs of active plate tectonics or continuously mobile lithospheres, appear to operate in the stagnant-lid mode. Indeed, considering the admittedly meager statistics of our own solar system, the stagnant lid might very well be the dominant mode of convection in terrestrial planets (e.g., see O'Rourke and Korenaga, 2012). However, the very presence of mobile plates on Earth shows that its lithosphere–mantle system has important effects that mitigate the demobilization of the top thermal boundary layer caused by temperature-dependent viscosity. In fact, it is not at all clear that the extreme Arrhenius-type temperature dependence of mantle or lithospheric viscosity actually occurs from a practical standpoint; that is, as will be discussed later (see **Sections 7.07.5.4** and **7.07.6.2**), there are multiple competing microscopic deformation mechanisms and the dominant one tends to be the weakest, that is, the one that relieves stress most efficiently.

7.07.3.6 A Brief Note on the Influence of Sphericity

Clearly because the Earth is primarily a sphere, models of convection in thick spherical fluid shells would appear to be most realistic (e.g., Bercovici et al., 1989a; Bunge et al., 1997; Choblet et al., 2007; Coltice et al., 2012; Foley and Becker, 2009; Tackley et al., 1993; van Heck and Tackley, 2008; Zhong et al., 2000). However, the effects of sphericity are possibly not so fundamental to the generation of plates. In particular, sphericity tends to break the symmetry between the top and the bottom boundary layers, and upwellings and downwellings in the opposite sense of what we think is relevant for the Earth. For purely basally heated spherical shells, conservation of energy prescribes that for a convective solution to be stable

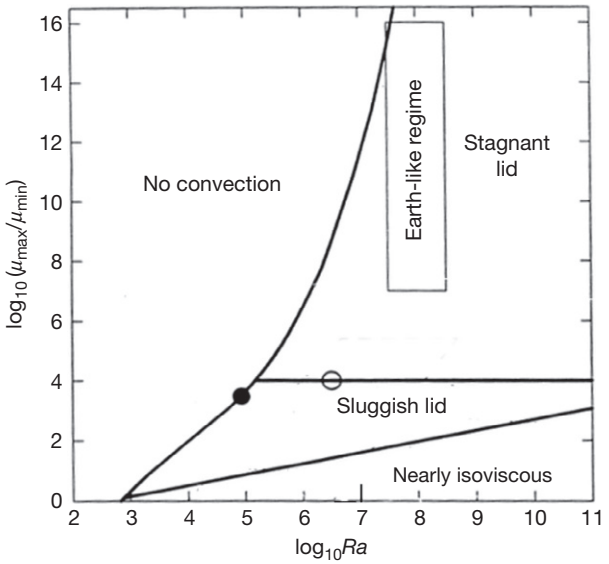


Figure 12 Diagram showing the different convective regimes in '*Ra* versus viscosity ratio' space for convection in fluid with temperature-dependent viscosity; μ_{max} and μ_{min} are the maximum and minimum allowable viscosities of the fluid, respectively. See text for discussion. Adapted from Solomatov V (1995) Scaling of temperature dependent and stress dependent viscosity convection. *Physics of Fluids* 7: 266–274.

(i.e., steady or statistically steady), the total heat input through the bottom boundary must be equal to the heat output through the top. However, because of sphericity, the bottom boundary has approximately four times less surface area than the top boundary through which heat passes; to compensate for this smaller area, the bottom thermal boundary layer generally has a larger temperature drop (and thus larger thermal gradients) across it than does the top boundary layer. This asymmetry leads to upwellings with larger temperature anomalies and velocities than the downwellings, which is opposite to the asymmetry between upwellings and downwellings thought to exist in the Earth. Thus, effects like internal heating and temperature-dependent viscosity are even more important in order to overcome the asymmetry imposed by sphericity and to give a more Earth-like asymmetry.

7.07.3.7 Poloidal and Toroidal Flow

Convective motion, with its upwelling and downwelling currents, and the associated divergent and convergent zones at the surface and lower boundary, is also called *poloidal* flow. Basic convection in highly viscous fluids essentially has only poloidal flow. As already discussed in Section 7.07.2.1.1, a great deal of the Earth's plate motion is indeed poloidal (i.e., the motion associated with spreading centers and subduction zones), but much of it also involves strike-slip motion and spin of plates, which is, again, *toroidal* motion (Cadek and Ricard, 1992; Dumoulin et al., 1998; Hager and O'Connell, 1978, 1979, 1981; Kaula, 1980; Lithgow-Bertelloni et al., 1993; O'Connell et al., 1991). The existence of toroidal flow in the Earth's plate-tectonic style of mantle convection is a major quandary for geodynamicists and is at the heart of a unified theory of mantle convection and plate tectonics. Therefore, the reason that toroidal motion does not exist in basic convection deserves some explanation.

Most models of mantle convection treat the mantle as nearly incompressible, that is, as if it has constant density. In fact, convection models must allow for thermal buoyancy; thus, they really treat the mantle as a *Boussinesq* fluid; this means that while density is a function of temperature (and thus actually not constant), the density fluctuations are so small that the fluid is still essentially incompressible *except* when the density fluctuations are acted on by gravity (see Chapters 7.02 and 7.04). Thus, while the fluid behaves as if incompressible, it can still be driven by buoyancy. Moreover, even without thermal density anomalies, the mantle is still not really incompressible; due to increase in pressure, its density changes by almost a factor of 2 from the top to the bottom of the mantle (e.g., from the IASP91 reference Earth model, Kennet and Engdahl, 1991). However, mantle flow occurs on a much slower timescale than compressional phenomena – in particular acoustic waves – and thus, the mantle is really considered *anelastic* for which the separation of flow into poloidal and toroidal parts still applies (Bercovici et al., 1992; Glatzmaier, 1988; Jarvis and McKenzie, 1980; Tackley, 1996a). However, even the influence of this anelastic component of compressibility is rather small compared to other effects in thermal convection (Bercovici et al., 1992; Bunge et al., 1997; Tackley, 1996a) (see Chapter 7.02).

The incompressibility or Boussinesq condition requires that the rate at which mass is injected into a fixed volume must equal the rate at which the mass is ejected, since no mass can be compressed into the volume if it is incompressible; that is, what goes in must equal what goes out. Mathematically, when considering infinitesimal fixed volumes – the continuum mechanical equivalent, of a point in space – this condition is called the continuity equation and is written as

$$\nabla \cdot \mathbf{v} = 0 \qquad [2]$$

where $\mathbf{v}$ is the velocity vector and ∇ is the gradient operator. This equation says that the net divergence of fluid away from or toward a point is zero; that is, if some fluid diverges away from the point, an equal amount must converge into it in order to make up for the loss of fluid. The most general velocity field that automatically satisfies this equation has the form (in Cartesian coordinates)

$$\mathbf{v} = \nabla \times \nabla \times (\phi \hat{\mathbf{z}}) + \nabla \times (\psi \hat{\mathbf{z}}) \qquad [3]$$

(where $\hat{\mathbf{z}}$ is the unit vector in the vertical direction) since $\nabla \cdot \nabla \times$ of any vector is zero. The velocity field in eqn [3] uses only two independent quantities, namely, ϕ and ψ, to account for three independent velocity components v_x, v_y, and v_z; this is allowed because eqn [2] enforces a dependence of one of the velocity components on the other two, and thus, there are really only two independent quantities. The quantity ϕ is called the poloidal potential that represents upwellings, downwellings, and surface divergence and convergence and (as we will show later) is typical of convective motion. The variable ψ is the toroidal potential and involves horizontal rotational or vortex-type motion, such as strike-slip motion and spin about a vertical axis; toroidal flow is not typical of highly viscous convective motion (Figure 13).

What drives flow in a highly viscous fluid (where fluid particles are always at terminal velocity, i.e., acceleration is negligible) is the balance between buoyancy (i.e., gravitational) forces, pressure gradients, and viscous resistance; in simple convection with a constant-viscosity fluid, this force balance is expressed as

$$0 = -\nabla P + \mu\nabla^2\mathbf{v} - \rho g\hat{\mathbf{z}} \qquad [4]$$

where P is pressure, μ is viscosity, ρ is density, and g is gravitational attraction (on Earth's surface, $g = 9.8\ \mathrm{m\ s^{-2}}$). The viscous resistance term, proportional to μ in eqn [4], is due to an imbalance or gradients in stress, while stress is due to gradients in velocity – that is, relative motion between particles leading to shearing, stretching, and squeezing – imposed on a fluid with a certain stiffness or viscosity (see Chapter 7.02 this volume). Taking $\hat{\mathbf{z}}\cdot\nabla\times\nabla\times$ of eqn [4] eventually leads to

$$\mu\nabla^4\phi = -\rho' g \qquad [5]$$

where ρ' is the density anomaly (e.g., due to horizontal variations in temperature). Equation [5] shows that the *poloidal* motion (left side of the equation) is directly driven by buoyancy forces (right side). However, if we take $\hat{\mathbf{z}}\cdot\nabla\times$ of eqn [4], we only obtain

$$\nabla^2\psi = 0 \qquad [6]$$

which shows that there are no internal driving forces for toroidal flow. (Note that in deriving both eqns [5] and [6], we have removed a common operator $\nabla_H^2 = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$, which is the 2-D horizontal Laplacian, assuming the domain is horizontally uniform or periodic.) Thus, for isoviscous convection, toroidal motion does not occur naturally and the only way to generate it is to excite it from the top and/or the bottom boundaries; otherwise, toroidal motion does not exist. These same arguments hold for spherical geometry (e.g., see Busse, 1975; Bercovici et al., 1989b; Chandrasekhar, 1961), as well as for a compressible anelastic fluid (Bercovici et al., 1992; Glatzmaier, 1988; Jarvis and McKenzie, 1980), and even if the gravity g and viscosity μ are functions of height z (see Section 7.07.5.7.2).

Convective flow in a highly viscous medium can only drive toroidal motion by having platelike boundary conditions (Gable et al., 1991; Hager and O'Connell, 1978; Ricard and Vigny, 1989) or if there is a forcing term on the right side of eqn [6] related to the buoyancy or poloidal motion; as will be discussed later (Section 7.07.5.7.2), this forcing term only occurs if the viscosity varies laterally, that is, $\mu = \mu(x, y)$ (see Figure 14 for a qualitative explanation). The problem of how to get buoyancy-driven poloidal motion to induce toroidal flow, either through boundary conditions or through horizontal viscosity variations, is called the *poloidal–toroidal coupling* problem. This will be discussed in more detail in the succeeding text (Section 7.07.5.7).

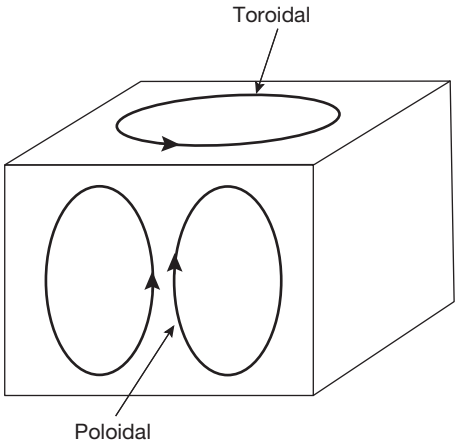


Figure 13 Cartoon illustrating simple flow lines associated with toroidal and poloidal motion.

7.07.4 Where Does Basic Convection Theory Succeed in Explaining Plate Tectonics?

Given some of the essential aspects of simple viscous convection reviewed earlier, we can examine those features of plate tectonics that are characteristic of basic convective flow. Subsequently (Section 7.07.5), we discuss those aspects of plate tectonics that are *poorly* explained by basic convection.

7.07.4.1 Convective Forces and Plate Driving Forces

7.07.4.1.1 Slab pull and convective downwellings

If we are to draw any analogy between simple viscous convection and platelike mantle flow, then we must identify subducting slabs with the downwelling of a cold upper thermal boundary layer. As discussed earlier in Section 7.07.2.1.2, the

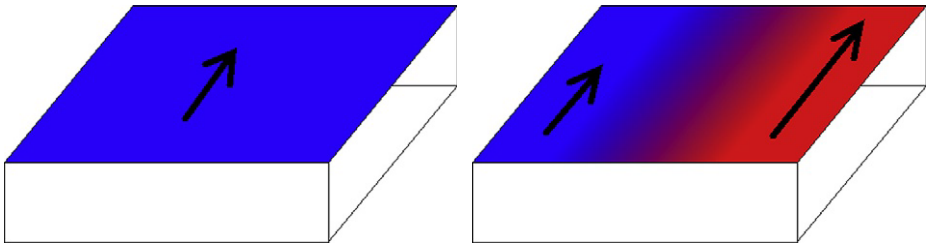


Figure 14 Cartoon illustrating the need for variable viscosity to obtain toroidal flow. A floating viscous fluid is drawn down a linear drain like a subduction zone. If the layer is uniform in viscosity, then the flow is uniformly into the drain (left). But if the viscosity varies laterally, especially with a gradient perpendicular to the flow direction (right), then the more compliant fluid will be drawn into the drain faster than the stiffer regions. The relative motion between the compliant and the stiff regions establishes a strike-slip shear-type flow with a finite toroidal velocity field.

correlation between the connectivity of a plate to a slab (i.e., the percent of its perimeter taken by subduction zones) and the plate's velocity argues rather conclusively for the dominance of slab pull as a plate driving force (see **Figure 4**). Therefore, if slabs are simply cold downwellings, then the descent velocity of a downwelling in simple convection should be comparable to that of a slab and by inference the velocity of a plate (in particular a fast slab-connected or *active* plate). (As mentioned earlier (**Section 7.07.3.2**), the pull of a slab on a plate is in fact a horizontal pressure gradient acting in the cold upper thermal boundary layer and caused by the low pressure associated with a slab pulling away from the surface; invariably, the pressure gradient is established so that the boundary layer or plate feeds the slab steadily and thus leads to the appearance that the slab is pulling the plate.) We can estimate the force and velocity of such a cold downwelling with simple scaling analysis and compare the calculated velocity with tectonic plate velocities.

Generalizing the scaling analysis and boundary layer theory of Davies and Richards (1992), we consider a cold top thermal boundary layer that is L long (from its creation at a divergent zone to its destruction at a convergent zone) and W wide (**Figure 15**). As it moves from the divergent to the convergent zone, the boundary layer thickens by vertical heat loss, which, in the horizontally moving boundary layer, is only due to thermal diffusion.

To estimate the boundary layer thickness $\delta(t)$, we apply dimensional homogeneity (e.g., see Bridgman, 1922; Furbish, 1997) (see **Chapter 7.04**). Assuming that only thermal diffusivity κ (with units of $m^2 s^{-1}$) and age t (where $t=0$ at the divergent zone where the boundary layer is initiated) control the growth of the boundary layer, then we write $\delta(t)=A\kappa^a t^b$ where A is a dimensionless constant (A is dimensionless since we assume no other dimensional properties of the system have any influence on the cooling process). The constants a and b are then determined to match the dimensions of either side of the equation for δ, leading to $a=b=1/2$. Indeed, a more careful analysis (see Turcotte and Schubert, 2002) shows that $\delta \approx \sqrt{\kappa t}$, within a factor of order unity.

When the boundary layer reaches the downwelling, its thickness is $\delta=\sqrt{\kappa L/v}$, where v is the as yet undetermined velocity of the boundary layer. After downwelling, the descending vertical current extends to depth D. Moreover, because the boundary layer involves a temperature contrast of ΔT (from the deeper fluid interior to the surface), then the layer and the vertical current it feeds have an average thermal anomaly of $\Delta T/2$. The buoyancy force of the downwelling current is therefore

$$F_B=[\rho\alpha\Delta T/2][\delta WD]g \qquad [7]$$

where the first factor in the square brackets represents the density anomaly of the downwelling and the second factor in brackets is the volume of the downwelling. The drag force on the downwelling current is the viscous stress acting on the current (which, in this example, is approximately the viscosity times the strain rate of the fluid as it is sheared vertically between the downwelling and a stagnation point roughly halfway back to the divergent zone) times the area of the current on which the stress acts, that is,

$$F_D=\left[\mu\frac{v}{L/2}\right][DW]=2\mu vDW/L \qquad [8]$$

The force balance $F_B-F_D=0$ yields the downwelling velocity

$$v=\frac{\rho\alpha\Delta T\delta Lg}{4\mu} \qquad [9]$$

Substituting $\delta=\sqrt{\kappa L/v}$ and isolating v, we obtain

$$v=\left[\left(\frac{\rho\alpha\Delta Tg}{4\mu}\right)^2 L^3\kappa\right]^{1/3} \qquad [10]$$

As we have seen for most forms of basic convection, $L\approx D$ (i.e., convection cells are roughly unit aspect ratio), in which case [10] is equivalent to the classic scaling law $v\sim(\kappa/D)Ra^{2/3}$. Thus, using $D=3\times10^6$ m, $\kappa=10^{-6}$ m^2 s^{-1}, $\rho=4000$ kg m^{-3}, $\alpha=3\times10^{-5}$ K^{-1}, $\Delta T=1400$ K, $g=10$ m s^{-2}, and $\mu=10^{22}$ Pa s (this viscosity represents the effective whole mantle viscosity that is dominated by the lower mantle), we obtain a velocity of $v\approx10$ cm year^{-1}. (We can adjust less well-constrained parameters, e.g., use, within reason, a lower or higher viscosity or a value of $L>D$; regardless the velocity remains of the order of 10 cm year.) This velocity is indeed the correct order of magnitude for velocities of active (i.e., slab-connected) plates,

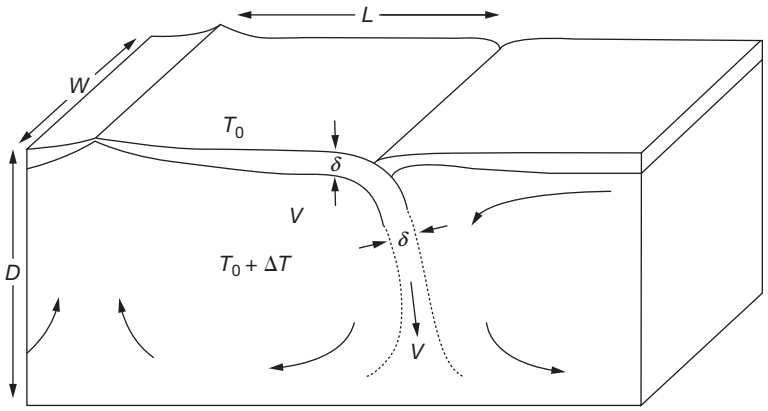


Figure 15 Cartoon illustrating the slab-pull force problem discussed in **Section 7.07.4.1.1**. For definition of labeled variables, refer to text (see **Section 7.07.4.1.1**). Adapted from Davies G and Richards M (1992) Mantle convection. *Journal of Geology* 100: 151–206.

strongly suggesting that the plate force deemed 'slab pull' is well modeled by a simple cold downwelling current in basic convection. Moreover, simple heat-flux arguments lead to an identical result (see Section 7.07.4.4), showing that slabs not only drive plate motions and mantle flow but also are the primary means for cooling off the mantle.

7.07.4.1.2 Ridge push is a convective lateral pressure gradient

The ridge-push plate force was initially conceived of as an edge force (i.e., acting on the edge of a plate at a mid-ocean ridge) due to the weight of a topographically high ridge (e.g., Forsyth and Uyeda, 1975). It has since been recognized as a force that is distributed across the area of the plate (Hager and O'Connell, 1981). Since plate area has little correlation with plate velocity (Forsyth and Uyeda, 1975), ridge push is typically assumed secondary to slab pull, or it is almost exactly canceled by mantle drag along the base of plates (Hager and O'Connell, 1981). Indeed, Lithgow-Bertelloni and Richards (1998) estimated that ridge push constitutes only 5–10% of the driving force due to subducted slabs. Nevertheless, after slab pull, ridge push is one of the more significant forces (and possibly the only driving force for plates lacking slabs, like the African and Antarctic Plates); thus, it deserves some discussion from the perspective of basic convection.

Ridges are topographically high because they are younger and composed of hotter more buoyant material than the surrounding material, which has cooled and grown heavier since moving away from the ridge. The subsidence of material away from a ridge is slow enough that the entire ridge–lithosphere system is essentially in isostatic balance, at least far from subduction zones. If topography is isostatic, then there is a *compensation depth* beneath the lithosphere at which every overlying, infinitely high, column of material with equal cross-sectional area weighs the same; this also means that hydrostatic pressure (i.e., weight per area) is the same at every horizontal position along and below the compensation level. However, columns that do not extend to the compensation depth do not weigh the same; in particular, the column centered on the ridge axis weighs the most since off-axis columns still include too much light material (like water) and not enough heavy material (like lithosphere) to equal the weight of the column of hot asthenosphere beneath the ridge. Thus, for depths less than the isostatic compensation depth, pressure is not horizontally uniform and is in fact highest beneath the ridge axis, thereby causing a pressure gradient pushing outward from the ridge (see Turcotte and Schubert, 2002); however, this pressure gradient exists everywhere within the lithosphere (i.e., wherever seafloor depth changes laterally) and not as a line force at the ridge axis.

How does this ridge-force pressure gradient relate to convective forces? As discussed previously (Section 7.07.3.2), the horizontal pressure gradient in a convective thermal boundary layer is what drives the lateral motion of the boundary layer. If the upper surface were deformable, the pressure highs at divergent zones would push up the surface, and the pressure lows at convergent zones would pull down on the surface. Thus, the pressure gradient in the thermal boundary layer would be manifested as surface subsidence from a divergent to convergent zone. In fact, as discussed in the next section, the resulting surface subsidence of a simple convective boundary layer is nearly identical to that observed for oceanic lithosphere. Thus, the ridge-push pressure gradient is essentially indistinguishable from the pressure gradient in a convective thermal boundary layer.

7.07.4.2 Structure of Ocean Basins

Large-scale variations in bathymetry, or seafloor topography, can be related to simple convective processes, in particular the cooling of the top thermal boundary layer. Although we have already discussed the cooling-boundary-layer concept in the previous section, we will reiterate the arguments in slightly more detail in order to examine the validity of the underlying assumptions.

As mentioned earlier (Section 7.07.3.2.2), when the top thermal boundary layer of a convection cell moves horizontally, it cools by heat loss to the colder surface. In simple convection, the surface is assumed impermeable, that is, no mass crosses it and thus, heat loss out of the top thermal boundary layer is assumed entirely due to thermal diffusion; that is, since the boundary layer abuts the impermeable surface, vertical heat loss due to ejection of hot mass or ingestion of cold mass across the surface is assumed negligible. While this assumption is reasonable for most of the extent of the thermal boundary layer, it causes convective theory to mispredict some details of ocean-basin topography and heat flow (see Section 7.07.5.2).

As discussed earlier in Section 7.07.3.2.2, the assumption of diffusive heat loss means that thickening of the boundary layer is controlled by thermal diffusivity κ, with units of $m^2\,s^{-1}$; thus, again by dimensional homogeneity (see Section 7.07.4.1.1), the boundary layer thickness $\delta \sim \sqrt{\kappa t}$. As the thermal boundary layer thickens, it weighs more and thereby pulls down on the surface with increasing force; if the surface is deformable, it will be deflected downward (until isostasy is established). The surface subsidence thus mirrors the increasing weight of the boundary layer; the weight increases only because δ grows, and thus, surface subsidence goes as $\sqrt{t}$. Therefore, seafloor subsidence is predicted by convective theory (more precisely convective boundary layer theory, Turcotte and Oxburgh, 1967) to increase as the square root of lithospheric age; this is called the $\sqrt{\text{age}}$ law. Moreover, this type of boundary layer theory predicts that the heat flow (units of $W\,m^{-2}$) out of the ocean floor should obey a $1/\sqrt{\text{age}}$ law, which can be inferred by the fact that the heat flow across the boundary layer is mostly conductive and thus goes as $k\Delta T/\delta$ (where k is thermal conductivity and ΔT is the temperature drop across the thickening thermal boundary layer, as also defined in Section 7.07.4.1.1). Profiles, perpendicular to the spreading center, of seafloor bathymetry and heat flow versus age (Figure 16) show that, to first order, seafloor subsidence and heat flow do indeed follow boundary layer theory, that is, obey $\sqrt{\text{age}}$ and $1/\sqrt{\text{age}}$ laws, respectively (Parsons and Sclater, 1977; Sclater et al., 1980; Stein and Stein, 1992; Turcotte and Schubert, 2002), further emphasizing that the oceanic lithosphere is primarily a convective thermal boundary layer.

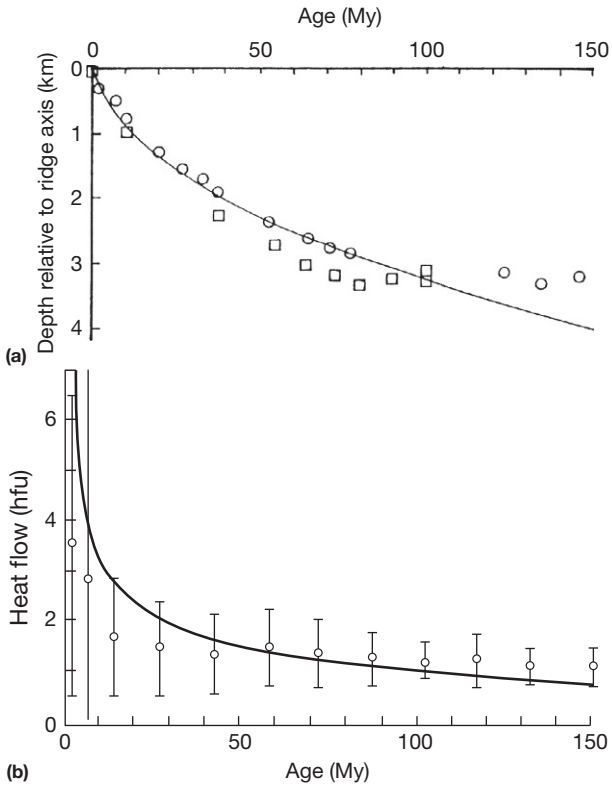


Figure 16 Bathymetry – that is, seafloor topography – (a) and heat flow (b) versus age for typical ocean floor. Bathymetry is for the North Pacific (circles) and North Atlantic (squares) oceans; heat flow (in hfu = 41.84 mW m^{-2}) shows average global values (circles) and associated error bars. Both (a) and (b) show the predictions from convective boundary layer theory (solid curves), that is, a $\sqrt{\text{age}}$ law for bathymetry and a $1/\sqrt{\text{age}}$ law for heat flow. The theoretical curves fit the observations reasonably well, except near and far from the ridge axis (cf. Parsons and Sclater (1977); Sclater et al. (1980); Stein and Stein, 1992). Adapted from Turcotte D and Schubert G (2002) *Geodynamics*. Cambridge: Cambridge University Press.

However, at a greater level of detail, the $\sqrt{\text{age}}$ laws fail; this will be discussed later (**Section 7.07.5.2**).

7.07.4.3 Slablike Downwellings Are Characteristic of 3-D Convection

In simple, purely basally heated plane-layer convection, the convective pattern is often characterized by interconnected sheets of downwellings. For example, the hexagonal pattern is characterized by one upwelling at the center of each hexagon, surrounded by a hexagonal arrangement of downwelling walls (**Figure 7**). A similar situation occurs for spherical systems (Bercovici et al., 1989b; Busse, 1975; Busse and Riahi, 1982, 1988; Zhong et al., 2000). When internal heating of the fluid is included, the upwellings and downwellings can be so imbalanced as to destroy any trace of symmetry. Even so, the downwellings can persist as linear, albeit no longer interconnected, sheetlike downwellings at low to moderate Rayleigh numbers (Bercovici et al., 1989a; Weinstein and Olson, 1990), and give way to downwelling cylinders and blobs at high Rayleigh numbers (Bunge et al., 1996; Glatzmaier et al., 1990; Sotin and Labrosse, 1999); however, even at high Ra, the linearity of the downwellings is recovered by allowing viscosity to increase with depth, as is thought to occur in the Earth (Bunge et al., 1996) (see **Figure 9**). Thus, although sheetlike downwellings do not always occur in simple models of thermal convection, they are a very common feature. In that sense, the occurrence of slablike downwellings in mantle convection is characteristic of basic thermal convection. However, as we will discuss later, there are many other characteristics of subducting slabs that are unlike simple convective flow (see **Chapter 7.09**).

7.07.4.3.1 Cylindrical upwellings and mantle plumes are also characteristic of 3-D convection

Although mantle plumes and hot spots are not, strictly speaking, part of plate tectonics, they play an important role in our ability to measure absolute plate motions (e.g., because relative positions between hot spots appear to be fixed (Morgan, 1972) or change more slowly than plate motions (Molnar and Stock, 1987)) and in understanding the nature of mantle convection and the relative sizes of its thermal boundary layers.

As with sheetlike downwellings, cylindrical or plumelike upwellings are prevalent in simple thermal convection. (It is possible to get different shapes – for example, sheetlike – of upwellings, but they tend to break down into more columnar features before reaching the surface.) However, such upwellings always require a significant bottom thermal boundary layer heated from below by a relatively hot reservoir, such as the Earth's outer core; this boundary layer provides a source region for plumes, and its thickness therefore determines their size (Bercovici and Kelly, 1997; Christensen, 1984c). In basic convection with significant internal heating, the bottom thermal boundary layer typically has a much smaller temperature drop than does the top thermal boundary layer. In this case, upwellings still retain some plumelike quality though they are fewer in number and/or weak, relative to downwellings or slabs. However, the strength of plumes may depend on the large-scale style of convection; in particular, mobile-lid or platelike convection allows for a cooler mantle than in stagnant-lid convection, thus a bigger bottom thermal boundary layer, hence more robust plumes (Jellinek and Manga, 2004).

Overall, the existence of cylindrical upwelling plumes in the mantle – indirectly inferred from hot spot volcanism and geochemical analyses (Allégre, 1982; Duncan and Richards, 1991; Hofmann, 1997; Zindler and Hart, 1986) and perhaps most directly in seismic analysis (Montelli et al., 2004; Nataf and VanDecar, 1993; Shen et al., 1998; Wolfe et al., 1997, 2009) – is consistent with basic models of thermal convection. However, even the weak plumes appearing in 3-D models of internally heated convection are an integral part of the convective flow; that is, while they may not have much influence on the dynamics of the top cold boundary layer, they typically position themselves in accordance with the whole convective pattern, typically as far from downwellings as possible (Weinstein and Olson, 1989). Plumes well integrated into the global convective circulation are very unlikely to establish themselves where they must pass through adverse conditions (e.g., beneath a downwelling or across a strong 'mantle wind'), and even if they do, they can only exist as evanescent features

(Labrosse, 2002; Steinberger and O'Connell, 1998) (see Chapter 7.10).

7.07.4.4 Slab Heat Flux and Plate Velocities

The Earth's mantle is thought to be about 70–80% heated internally by primordial heat and radiogenic sources, and the rest from other sources, primarily from cooling of the core (see Chapter 7.11). As plumes are thought to carry the heat from the core (see next section), heat flow through the oceanic lithosphere, or equivalently cooling by slabs, accounts for most of the Earth's heat flow. Indeed, one can use these heat flow arguments to predict slab and plate velocities and check for self-consistency with the boundary layer force balance arguments of Section 7.07.4.1.1, which thus verifies that slabs are an integral part of convective heat transport (Bercovici, 2003).

At a few 100 km depth beneath the lithosphere, mantle heat loss is primarily due to the downward injection of cold material by subducting slabs (analogous to dropping ice cubes in hot water). The energy-flux balance would thus require that $fQ = wA\rho c_p \Delta T$ where $Q \approx 30$–38 TW is the net heat output through the top of the mantle (Schubert et al., 2001) $f = 0.8$–0.9 is the fraction of Q accounted for by slab cooling (since the remaining heat transport by mantle plumes probably accounts for of order 10–20% of the net heat flux; Davies, 1988a; Jaupart et al., 2007; Lay et al., 2008; Pollack et al., 1993; Sleep, 1990), $\rho \approx 3000\ \mathrm{kg\,m^{-3}}$ is slab density, $c_p = 1000\ \mathrm{J\,kg\,K^{-1}}$ is heat capacity, $\Delta T \approx 700$ K is the average slab thermal anomaly as in Section 7.07.4.1.1, and w is a typical vertical velocity of a slab. Lastly, $A \approx 2\pi R\delta$ is the total horizontal cross-sectional area of all slabs crossing this particular depth where $\delta \approx 100$ km is a typical slab thickness, and the horizontal length of all slabs is estimated by the circumference of the Earth since most slabs occur in a nearly large circle around the Pacific basin and thus $R \approx 6000$ km. Using these numbers, we can solve for slab velocity $w = fQ/(\rho c_p \Delta T \delta 2\pi R) \approx 10\,\mathrm{cm\ year^{-1}}$, which is precisely the typical velocity for fast, active plates and predicted by slab-pull arguments (Section 7.07.4.1.1). In the end, that 'active' plates cool, thicken, subside, and eventually sink as subducting slabs at plate velocities, while also cooling the mantle is tantamount to saying slabs and plates are convective currents.

7.07.4.5 Mantle Heterogeneity and the History of Subduction

Up to this point, one can construct a simple view of the connection between plate tectonics and mantle convection in which subduction zones are associated with convective downwellings and plumes with upwellings; whether ridges are also associated with active convective upwellings is less certain, if not doubtful (see Section 7.07.5.6). If heating is largely internal, then downwelling of the upper boundary layer, or lithosphere, at subduction zones would then be expected to dominate the structure of mantle temperature and density heterogeneity. Indeed, this seems to be the case, but the picture is complicated by the fact that plate motions, and hence subduction zones, are ephemeral on timescales of tens to hundreds of million years. Thus, one does not expect a perfect correspondence between present-day subduction zones and mantle downwellings, and it is necessary to invoke plate motion history in addressing this issue.

Averaged over the past 120–150 My, the locations of subduction zones (in a hot spot reference frame) correspond remarkably well to the location of high-seismic-velocity anomalies (cold material) in the mantle, at least at very long wavelengths (Richards and Engebretson, 1992) (**Figure 17**), a view supported in some detail by high-resolution tomographic studies of mantle structure (Fukao et al., 1992; Grand et al., 1997; van der Hilst et al., 1997; see also Volume 1). In fact, both the geoid (the Earth's gravitational equipotential surface; see Ricard et al., 1993) and global plate motions (Lithgow-Bertelloni and Richards, 1998) are well predicted using a mantle density model derived from slabs subducted during Cenozoic and Mesozoic times. However, there appears to be little evidence in these analyses of deeply seated upwellings beneath mid-ocean ridges that would provide plate driving forces of similar magnitude, although the African swell is a possible exception to this inference (Lithgow-Bertelloni and Silver, 1998). On the whole, one can infer a strong relationship between the mass of slabs injected at plate velocities since the Mesozoic with current observations of tomographic density

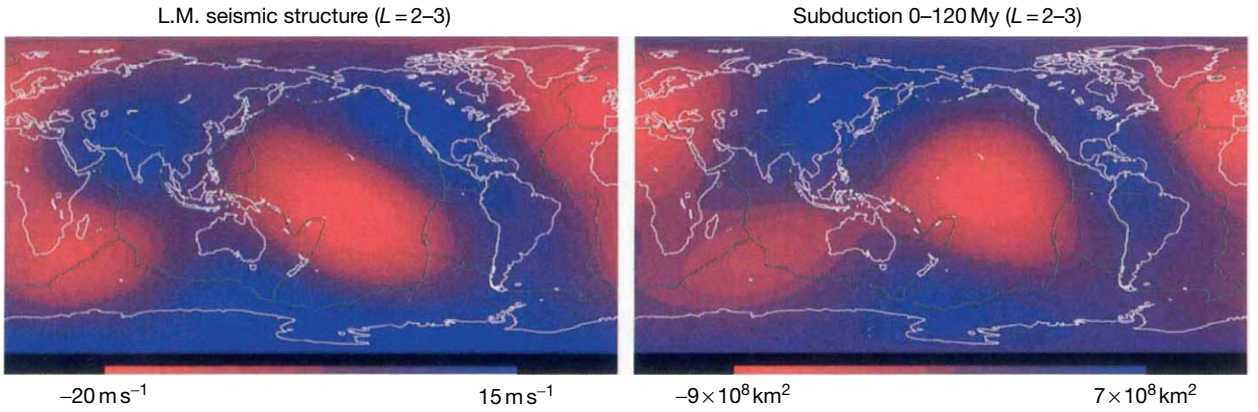


Figure 17 Comparison of long-wavelength lower-mantle tomography (left) with time-integrated volume flux of slabs over the last 120 My (right). Reproduced from Richards MA and Engebretson DC (1992) Large-scale mantle convection and the history of subduction. *Nature* 355(6359): 437–440.

heterogeneity, and geoid anomalies, which confirms the linkage between mantle convection and plate tectonics. However, this inference bears little on the question of how the plates themselves are generated in the system.

7.07.4.6 Temperature-Dependent Viscosity, Internal Resistive Boundaries, and the Aspect Ratio of Convection Cells

The Pacific Plate, with most of the kinetic energy and subduction zones of the plate-tectonic system, can in many ways be considered the top of the Earth's dominant convection cell. As such, it is a very large convection cell with an excessively large aspect ratio (much longer than it is possibly deep). (The size of several other massive plates could also be used to argue for large aspect ratio cells; however, except for the Indo-Australian plate – which in itself may be composed of several other plates (Gordon et al., 1998) – no other superlarge plate is considered *active*, i.e., slab-connected and 'fast,' and therefore does not really qualify as a possible convection cell.) Large aspect ratio or long-wavelength convection cells in the mantle are also inferred from seismic tomography (Su and Dziewonski, 1992) (see also **Figure 17**). Because of the dominance of such long-wavelength features, tomography and by inference convection are deemed to have a *reddened* spectrum.

Therefore, in many ways, mantle convection in the Earth is thought to be characterized by very large aspect ratio convection cells. As discussed previously (**Section 7.07.3.5**), convection with temperature-dependent viscosity can allow for large aspect ratio cells (e.g., Weinstein and Christensen, 1991), in particular in the *sluggish* convection regime (with moderately temperature-dependent viscosity and a sluggish but mobile upper thermal boundary layer; see Solomatov, 1995). This effect can theoretically allow for a Pacific-sized convection cell; however, if the Earth's mantle were only a temperature-dependent-viscosity fluid, estimates of its convective regime would place it in the *stagnant-lid* regime where the top thermal boundary layer is essentially frozen (Solomatov, 1995) and thus would be more characterized by unit aspect ratio convection cells.

However, other basic phenomena may also cause long-wavelength convection cells; these effects primarily involve an internal resistive boundary that, while not a feature of the fundamental fluid dynamics of convection, can be readily added to simple convection models to improve their correspondence with the Earth.

As is well known, the mantle is nominally divided into the upper mantle and the lower mantle at the 660 km seismic discontinuity (see review by Silver et al., 1988). This discontinuity is most likely a phase change from spinel to perovskite plus magnesiowiistite (Ito et al., 1990; Liu, 1979) and also involves a viscosity jump with depth by most likely a factor of 30–100 (e.g., Corrieu et al., 1994; Hager, 1984; King, 1995; Ricard et al., 1984, 1993; Richards and Hager, 1984). Although the nature of convective mass transfer across this 660 km boundary has been the subject of extensive debate, seismic tomography has given compelling evidence for transfer of both slabs (Creager and Jordan, 1984; Fukao et al., 1992; Grand, 1994; Grand et al., 1997; van der Hilst et al., 1991, 1997) and plumes (Shen et al., 1998; Wolfe et al., 2009) across the phase change. However, because of both the phase change and the viscosity jump, this boundary can still provide resistance to downwellings as they attempt to penetrate into the lower mantle. The phase change resists mass transfer because it is inferred to have a negative Clapeyron slope, which means that in relatively cold material, the phase transition will occur at a higher pressure than in hotter material. Thus, a cold downwelling moves the phase transition to greater pressures, causing a downward deflection of the phase boundary that acts to buoy the slab and resist its downward motion (Christensen and Yuen, 1985; Honda et al., 1993; Machetel and Weber, 1991; Schubert et al., 1975; Tackley et al., 1993, 1994; Weinstein, 1993). A viscosity jump retards the motion of downwellings because of the greater resistance of the material beneath the boundary. However, the viscosity jump need not be uniquely associated with the 660 km boundary; a low-viscosity asthenosphere or a smooth increase of viscosity with depth can induce similar effects (Ahmed and Lenardic, 2010; Lenardic et al., 2006). In any case, convection tends to develop large-volume downwellings – and thus large-wavelength convection cells – in order to gather enough negative buoyancy to penetrate more resistive lower layers. This 'reddening' of the convective spectrum (thus creating large aspect ratio, nearly Pacific size cells) has been well documented in 3-D models of spherical convection at reasonably high Rayleigh number for both the phase change (Tackley et al., 1993, 1994) and the viscosity-jump effects (Bunge et al., 1996; Tackley, 1996b).

In the end, a convecting mantle with either a moderately temperature-dependent viscosity or an internal resistive boundary can generate Earth-like, large-wavelength (or long aspect ratio) plate-sized convection cells. However, while these effects move us closer to obtaining platelike scales in convection models, none actually generate plates.

7.07.5 Where Does Basic Convection Theory Fail in Explaining Plate Tectonics?

Although models of basic thermal convection in the Earth's mantle can account for numerous features of plate tectonics, there remain many unresolved issues. Next, we survey some of the major apparent paradoxes and inconsistencies between convection and plate tectonics, as well as the progress in trying to solve them.

7.07.5.1 Plate Forces Not Well Explained by Basic Convection

While slab pull and to a large extent ridge push (really, distributed ridge push) are manifestations of simple convection, other forces are not. Most of the forces unaccounted for by convection are related to some feature of plate tectonics that convection does not (or does not easily) generate. One such force is transform resistance, as this involves strike-slip motion that cannot be readily generated, if generated at all, in simple convection. The lack of excitation of strike-slip motion is a major shortcoming of basic convection theory; however, we will defer discussion of this until later (see **Section 7.07.5.7**). The precise distribution of the *ridge push* force also cannot be predicted in simple convection models since they do not generate narrow passive divergent zones, that is, thin ridges fed by

shallow nonbuoyant upwellings. The lack of formation of passive but focused upwellings is also a significant shortcoming of basic convection theory and will be discussed later as well (see **Section 7.07.5.6**). Finally, collision resistance does not explicitly occur in basic thermal convection models, which do not include chemical segregation and continental formation. This phenomenon can, in effect, be treated by thermochemical models that allow melting and surface emplacement of chemically buoyant continental material, which resists being drawn into downwellings (Lee et al., 2005; Lenardic and Kaula, 1996).

7.07.5.2 Structure of Ocean Basins (Deviations from the $\sqrt{\text{age}}$ Law)

As discussed earlier (**Section 7.07.4.2**), the classic bathymetry and heat flow profiles (**Figure 16**) show a predominant $\sqrt{\text{age}}$ dependence. However, bathymetry and especially heat flow deviate significantly from the $\sqrt{\text{age}}$ curve near the ridge axis itself and far from the ridge axis. In particular, the $\sqrt{\text{age}}$ prediction assumes that only diffusive transport of heat occurs in the upper thermal boundary layer, which likely becomes invalid both at and far from the ridge axis. At the ridge itself, heat transport by magma migration (e.g., see Katz, 2008) and hydrothermal circulation (Lister, 1972; Stein and Stein, 1992) are highly significant and likely account for suppression of axial bathymetry and heat flow profiles. Incorporation of realistic or more complete transport phenomena relevant to ridges (e.g., magmatism and/or water ingestion) into convection models is also related to the problem of how to generate focused but passive ridges (see **Section 7.07.5.6**).

The flattening of bathymetry, relative to the $\sqrt{\text{age}}$ law, far from the ridge axis has had various explanations. It has been proposed that flattening is the result of extra heating of the lithosphere (causing effective rejuvenation) brought on by (a) secondary small-scale convection and/or internal heating (Huang and Zhong, 2005; Parsons and McKenzie, 1978), (b) viscous heating (Schubert et al., 1976), and (c) mantle plumes (Davies, 1988b); nonthermal effects, such as due to high pressure associated with constricted asthenospheric flow, have also been proposed (Morgan and Smith, 1992). However, since 3-D convection itself differs significantly from the 2-D boundary layer theory from which the $\sqrt{\text{age}}$ law is derived (Turcotte and Oxburgh, 1967), such flattening and other deviations in the far-axis bathymetry are perhaps not altogether surprising from a mantle convection perspective.

7.07.5.3 Changes in Plate Motion

Plate motions evolve with various timescales. Some are clearly related to mantle convection, such as those associated with the Wilson cycle (Wilson, 1966), that is, the periodic formation and breakup of Pangaea, approximately every 500 My. Various hypotheses have been proposed to explain the dispersal of supercontinents. The conventional view is that supercontinents provide an insulating blanket, and thus, with radiogenic heating, they warm the underlying mantle, eventually inducing a hot upwelling that weakens and breaks up the overlying lithosphere (Coltice et al., 2007, 2009; Gurnis, 1988; Lowman and Jarvis, 1996). However, this mechanism requires more radiogenic heating than may actually exist (Korenaga, 2008). In the same vein, the present-day continents do not suggest that they stand above hotter than normal mantle; indeed, their basal heat flux appears very low (e.g., Guillou et al., 1994). However, this recent observation might not be generally relevant; in particular, Rolf et al. (2012) showed that while small subcontinental temperature anomalies occur as continents are drifting, they can become significantly larger when continents assemble into a supercontinent. Regardless, whether supercontinent insulation induces sufficient heating to induce dispersal remains a matter of debate (see Bercovici and Long, 2014; Heron and Lowman, 2011; Lenardic et al., 2005, 2011).

Plate motion changes that have occurred on shorter timescales are even more difficult to understand. The plate-tectonic history recorded in paleomagnetic data and hot spot tracks consists of long stages of quasisteady motions separated by abrupt reorganizations.

Stages of quasisteady motion are reasonably well explained in terms of plate forces (slab pull, ridge push, mantle drag, etc; Forsyth and Uyeda, 1975) or equivalently convective buoyancy from large-scale heterogeneities (e.g., Lithgow-Bertelloni and Richards, 1995; Ricard et al., 1989). Gradual reversals can also be explained by convective motion; for example, the notion of thermal blanketing can be extended to allow for plate reversals

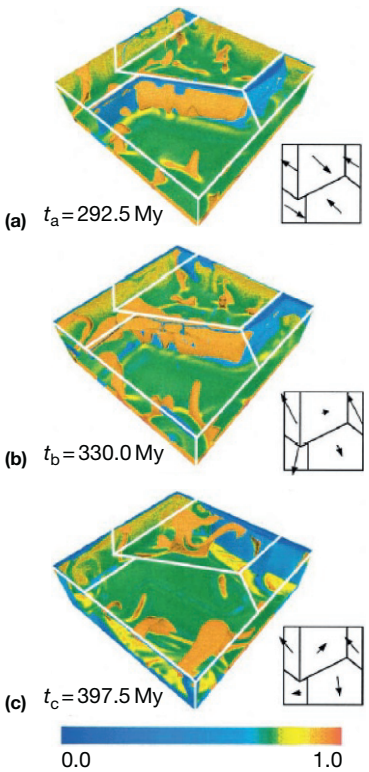


Figure 18 The convection model of King et al. (2002) shows that reversals in plate motion can occur when converging flow over a cold downwelling (a) draws in hot subcontinental mantle (b) that annihilates the downwelling, subsequently changing the polarity of the convergent margin to a divergent one (c). Adapted from King SD, Lowman JP, and Gable CW (2002) Episodic tectonic plate reorganizations driven by mantle convection. *Earth and Planetary Science Letters* 203(1): 83–91.

in that hot subcontinental mantle anomalies can effectively annihilate slabs and thus cause radical changes in plate motion (King et al., 2002; Lowman et al., 2003) (see **Figure 18**).

Abrupt changes in plate motion, however, are not easily related to convective processes. The most dramatic plate motion change is recorded in the Hawaiian–Emperor bend, dated at 47 Ma (Sharp and Clague, 2006; Wessel and Kroenke, 2008); this bend suggests a velocity change of a major plate of ~45° during a period no longer than 5 My, as inferred from the sharpness of the bend. Convective plate driving forces, such as due to sinking slabs, cannot change directions much faster than the time to lose or erase a thermal anomaly, which is of order several tens of millions of years, based on descent at a typical convective velocity (see **Section 7.07.4.1.1**). Therefore, changing convective motions in less than ~5 My is physically implausible. Abrupt changes may be due to nonconvective sources such as rapid rheological response and fast adjustments in plate boundary geometries, such as due to fracture and rift propagation (e.g., Hey and Wilson, 1982; Hey et al., 1995). Plate reorganizations due to (a) the annihilation of a subduction boundary by rapid slab detachment and/or continental collision (e.g., Bercovici et al., 2015), (b) loss of a ridge and/or trench by subduction of a ridge (e.g., Thorkelson, 1996), or (c) the initiation of a new subduction zone (e.g., Hilde et al., 1977) possibly occur on relatively rapid timescales, although the timing of such mechanisms is not well constrained (Richards and Lithgow-Bertelloni, 1996). Offsets in plate age and thickness along already weak oceanic transform boundaries may provide excellent sites for the initiation of subduction (e.g., Hall et al., 2003; Stern, 2004; Stern and Bloomer, 1992; Toth and Gurnis, 1998) yielding a possible mechanism for plate motion changes; this, coupled with the notion of transform faults as long-lived weak 'motion guides' (see **Section 7.07.6.1.2**), emphasizes the importance of transform boundaries in the plate–mantle system. The possible mechanism for abrupt plate motion changes underscores the need to understand the interactions between the long timescale convective processes and the short timescale effects associated with the lithosphere's rheological response, for example, by fracture, fault sliding, and strain localization (see **Section 7.07.6**).

7.07.5.4 Platelike Strength Distributions: Weak Boundaries and Strong Interiors

One of the platelike characteristics of the Earth's top thermal boundary layer is its strength or viscosity distribution. Most of the boundary layer is strong or of high viscosity; on Earth, these strong regions are separated by narrow zones of intense deformation (i.e., the plate margins), which are necessarily weak to permit the observed levels of strain. The cause for such a strength distribution is at the heart of understanding how plate tectonics arises from convection.

7.07.5.4.1 Do plates form because the lithosphere breaks?

It is often assumed that plates and their strength distribution form because the entire oceanic lithosphere is brittle and breaks under convective stresses. Thus, once the lithosphere is broken, it is not easily mended; the broken edges remain – more or less permanently – as sites of weakness and low-

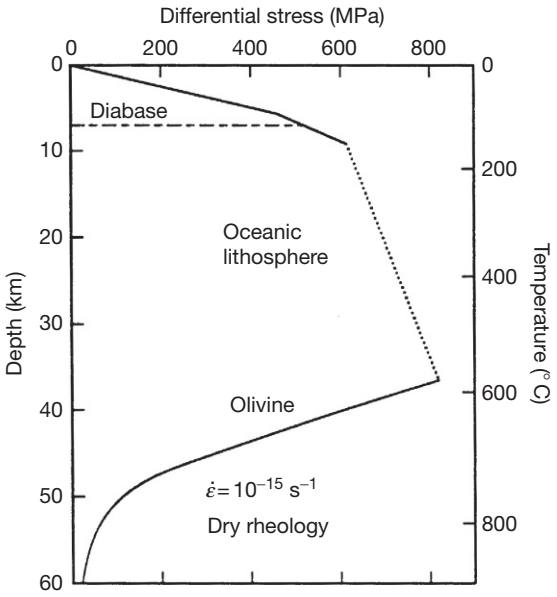


Figure 19 Strength of oceanic lithosphere (in terms of differential stress necessary to cause failure) versus depth and temperature for a model lithosphere of ~60 My age. The top straight segment represents brittle failure; the central dashed segment is combined brittle–ductile behavior (the strength values of which are not well constrained; hence, the dashed line should not be interpreted literally), and the lower curved segment represents the transition to fully ductile behavior (with no strength). Adapted from Kohlstedt D, Evans B, and Mackwell S (1995) Strength of the lithosphere: Constraints imposed by laboratory experiments. *Journal of Geophysical Research* 100: 17587–17602.

frictional sliding (e.g., Davies and Richards, 1992). However, brittle failure cannot account for weakening over most of the oceanic lithosphere. Indeed, brittle failure and frictional sliding cease at shallow depths, on the order of 10 km (Kohlstedt et al., 1995), and give way to semiductile/semibrittle behavior that involves ductile cracking with noncontiguous distributions of microcracks (Evans and Kohlstedt, 1995; Kohlstedt et al., 1995). This semiductile behavior in fact exists over greater depths than does frictional sliding (**Figure 19**). Moreover, grain-size reduction in shear zones (through mechanisms like dynamic recrystallization) is thought to allow weakening and localization in the middle to deep lithosphere, as evident in peridotitic mylonites, that is, exposed mantle shear zones (Furusho and Kanagawa, 1999; Jin et al., 1998; White et al., 1980) (see also **Section 7.07.6.2.3**). The model of the broken lithosphere is therefore not valid. The lithospheric deformation mechanisms leading to plates are more likely controlled by other complex processes; this will be discussed further in **Section 7.07.6.2**.

7.07.5.4.2 Plateness

Given the considerations mentioned earlier, many convection models seeking to generate plates employ a continuum approach; that is, they do not break the top cold boundary layer into pieces, but maintain its continuity while employing rheologies that allow parts of it to become weaker than other parts. In doing so, they try to obtain the property of a platelike strength (or viscosity) distribution for their top thermal boundary layer. This property is quantifiable and has been called

plateness (Weinstein and Olson, 1992). A boundary layer with broad, strong, slowly deforming regions (representing strong plate interiors) separated by narrow zones of weak, rapidly deforming material (representing plate boundaries) have high plateness; clearly, convection in the Earth has a top thermal boundary layer with high plateness. A boundary layer with no strength variations, as occurs in basic convection with constant or only depth-dependent viscosity, has zero plateness.

Since plateness involves lateral variations in strength or viscosity, convection models can obviously only generate sufficient plateness if they employ a fully variable-viscosity law. Generally, viscosity is variable by virtue of being a function of the thermodynamic state (e.g., temperature and pressure), state of stress, and/or composition of the medium.

If viscosity is temperature-dependent, the top boundary layer will tend to be weakest at the divergent zone (where the boundary layer is thinnest), which approaches platelike behavior. However, the viscosity will be highest over the cold convergent zone, tending to immobilize it and eventually the entire boundary layer, which is of course one of the root causes for the subduction initiation problem (see **Section 7.07.5.5**); in the end, this also leads to a thermal boundary layer with little plateness. Therefore, it is a reasonable assumption that platelike strength distributions – or high plateness – require a lithospheric rheology that permits boundaries to be weak, independent of temperature.

At high stresses, mantle rocks undergo non-Newtonian dislocation creep wherein deformation responds nonlinearly to the applied stress; that is, the stress–strain rate constitutive law is nonlinear through a power-law relation such that

$$\dot{e} \sim \sigma^n \qquad [11]$$

where $\dot{e}$ is strain rate, σ is stress, and n is the *power-law index* (see **Figure 23**); for mantle rocks, $n=3$, typically (Evans and Kohlstedt, 1995; Weertman and Weertman, 1975). (Note that the mathematically correct constitutive law relates the strain-rate and stress tensors, and thus, strictly speaking, $\dot{e}$ and σ are the square roots of the second scalar invariants of their respective tensors; see Malvern, 1969.) The effective viscosity is

$$\mu \sim \sigma/\dot{e} \sim \dot{e}^{1/n-1} \qquad [12]$$

which means that for $n>1$, viscosity decreases with increasing strain rate. The power-law rheology thus causes regions of the material that deform rapidly to become softer, while slowly deforming regions become relatively stiff. (Comparable effects can be caused with other rheologies such as Bingham plastics and biviscous laws; however, these are just further mathematical models for essentially the same strain-softening effect represented by a power-law rheology.) However, when such non-Newtonian rheologies with $n=3$ are incorporated into 2-D convection models, they yield very little platelike behavior (e.g., Christensen, 1984a; Parmentier et al., 1976).

Other models, concentrating on plate formation, placed a thin non-Newtonian 'lithospheric' fluid layer atop a thicker convecting layer or mathematically confined the non-Newtonian behavior to a near-surface layer or model lithosphere (Moresi and Solomatov, 1998; Richards et al., 2001; Schmeling and Jacoby, 1981; Weinstein, 1996; Weinstein and Olson, 1992); see **Figure 20**. In these models, the non-Newtonian effect does indeed yield an upper layer with reasonably high *plateness*. As noted by King et al. (1991), 2-D models with non-Newtonian lithospheres are almost indistinguishable from those with imposed plate geometries (e.g., Davies, 1988c; Olson and Corcos, 1980) and lithospheric weak zones (e.g., King and Hager, 1990; Zhong and Gurnis, 1995a), provided the non-Newtonian models employ a power-law index n considerably higher than experimentally determined, that is, with $n>7$. Moreover, the simple power-law model even with large n is less than successful in 3-D models where there are more than two types of boundaries (see succeeding text). Indeed, platelike behavior is more than simply generating weak boundaries, since each type of boundary, that is, convergent, divergent, and strike slip, develops under very different deformational and thermal environments. Obtaining sufficient platelike behavior is invariably related to the nature of how the different boundaries form. As

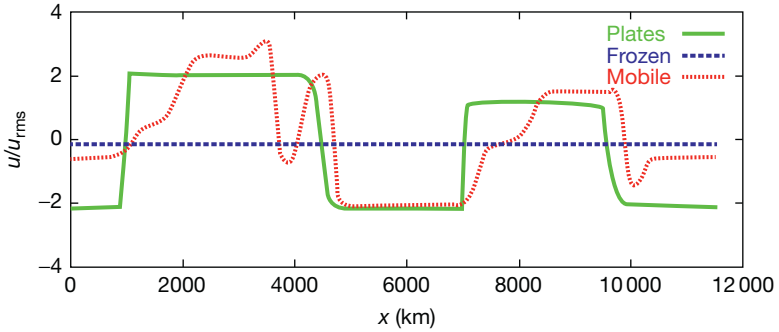


Figure 20 Surface velocity above two-dimensional simulations of convection with a variable viscosity top boundary. A mobile top boundary (red) occurs when the fluid is entirely or nearly isoviscous such that the top thermal boundary layer acts viscously and deforms continuously and smoothly. The 'frozen' top boundary (blue) occurs when the fluid viscosity is strongly temperature-dependent and the cold top boundary layer becomes stiff and immobile. The platelike top boundary (green) occurs when the fluid has a strongly non-Newtonian or plastic rheology; here, the surface velocity is composed of nearly blocklike structures of almost constant velocity, separated by narrow zones of intense deformation. The mobile boundary (red) is defined to have 'plateness' near zero, while the platelike one (green) has 'plateness' near unity (Weinstein and Olson, 1992). Adapted from Richards M, Yang W-S, Baumgardner J, and Bunge H-P (2001) Role of a low-viscosity zone in stabilizing plate tectonics: Implications for comparative terrestrial planetology. *Geochemistry, Geophysics, Geosystems* 2: 1026. http://dx.doi.org/10.1029/2000GC000115.

each type of boundary is uniquely enigmatic, they warrant individual discussion and thus will be the focus of the following sections.

7.07.5.5 Convergent Margins and Subduction Zones

Cold subducting slabs sinking into the mantle are, as discussed earlier, a clear manifestation of thermal convection, in terms of both driving circulation and plate motion (**Section 7.07.4.1.1**) and cooling the mantle (**Section 7.07.4.4**). However, subduction zones remain a highly enigmatic feature of platelike convection, and we review some of the more outstanding questions later. Many of these topics are covered in **Chapter 7.09** of this volume, thus, we will only survey them briefly within the context of plate–mantle interactions and plate generation.

7.07.5.5.1 Subduction initiation

How subduction zones initiate on Earth, but probably not the other terrestrial planets (the existence of subduction zones on Venus is, however, still arguable e.g., Schubert and Sandwell, 1995) is one of the most challenging issues in geodynamics (King, 2007; Mueller and Phillips, 1991; Stern, 2004). The fundamental difficulty with subduction initiation is that given mantle silicate's temperature-dependent viscosity, all terrestrial planets should be in a stagnant-lid regime of convection, that is, the extremely high viscosities of terrestrial lithospheres should prohibit them from being recycled into their underlying mantles, at least not on geologic timescales (**Section 7.07.3.5**).

The generation of weak zones on Earth that permit subduction of cold strong lithosphere is therefore a critical issue. Initiation and weakening mechanisms invoke, for example, rifting (Kemp and Stevenson, 1996; Nikolaeva et al., 2010; Schubert and Zhang, 1997) or sediment loading (Regenauer-Lieb et al., 2001) at passive margins and reactivation of preexisting fault zones from transform or spreading environments (Gurnis et al., 2004; Hall et al., 2003; Toth and Gurnis, 1998). All of these mechanisms have some observational motivation, although fault reactivation might be the most compelling (e.g., Lebrun et al., 2003). Nevertheless, each presupposes a rheological softening or shear-localizing mechanism that promotes lithospheric weakness specific to Earth; whether Earth's unique conditions involve the presence of liquid water, continental crust or cool enough surface temperatures remains an active area of debate.

7.07.5.5.2 Asymmetry of subduction

The cause for asymmetrical downwellings, that is, one-sided subduction, is also an important area of research for which much progress has occurred in recent years. The downwelling currents in simple convection are largely symmetric; that is, one downwelling is composed of two cold thermal boundary layers converging on each other (**Figure 9**). In the Earth, the downwellings associated with plate motions – that is, subducting slabs – are highly asymmetrical, that is, only one side of the convergent zone, only one plate, subducts. A common explanation for the asymmetry is that it is caused by chemically light continents counteracting the negative buoyancy of one of the boundary layers approaching the convergent zone (e.g., see Lenardic and Kaula, 1996); however, this hypothesis is not universally valid since there are a significant number of purely oceanic subduction zones.

The asymmetry may also reflect an inequality of pressure on either side of the subducting slab if it deviates slightly out of vertical; that is, a more acute corner flow on one side of the subduction zone induces lower pressure than on the opposite side, thereby causing the slab to be torqued to the more acute side, that is, to enhance its obliquity (see Turcotte and Schubert, 2002). The oblique downwelling may then act to impede the motion and descent of the boundary layer approaching the acute angle, since it would have considerable resistance to making a $>90^\circ$ downward turn, eventually leading to asymmetrical convergence. This effect is most significant for slabs that are effectively rigid relative to the surrounding mantle, in order that they act as stiff paddles while being lifted up; thus, if it occurs, the effect would theoretically be visible in basic convection calculations with strongly temperature-dependent viscosity. However, if the viscosity is so temperature-dependent that it induces such strong slabs, then it will also likely place convection itself into the *stagnant*-lid regime; that is, to make the slabs strong enough will also lock up the lithosphere (see **Sections 7.07.3.5** and **7.07.4.6**). To then adjust convection models to mitigate this extra problem requires proper initiation of subduction from cold, thick, and strong lithosphere (see **Section 7.07.5.5.1**).

Models with imposed weak zones in the lithosphere show that cold rigid lithosphere subducts readily and can even assume a fairly oblique slablike angle (Gurnis and Hager, 1988; Nakakuki et al., 2008; Zhong and Gurnis, 1995a), although both sides of the convergent margin usually undergo downwelling. Recent work has shown that permitting free surface deflection of the moving lithosphere as it bends into the subduction zone relieves much of the stress at this corner and promotes one-sided subduction (Crameri et al., 2012, results of which are displayed on the volume cover). One-sided subduction is further stabilized against eventual entrainment of the overriding lithosphere by including slab dehydration and subsequent mantle wedge hydration and melting (Gerya et al., 2008), which lubricates the acute corner flow adjacent to the slab (**Figure 21**).

7.07.5.5.3 Age of subduction zones

As discussed in **Section 7.07.2.1.3**, a significant unresolved enigma concerns the age of subduction zones. In particular, while the convective picture of plate motions would have plates subduct when they get old, cold, and heavy, in fact, subduction zones are distributed in age from nearly 0 (i.e., subducting ridges) to roughly 200 My (see Becker et al., 2009; Coltice et al., 2012). One possible explanation is that dehydration-induced stiffening of the lithosphere occurs during ridge partial melting and is thus age-independent (i.e., it does not get any stiffer with time); subduction is thus age-independent if resistance to subduction is just from bending this strong lithosphere, which can occur any time there is sufficient negative buoyancy (Conrad and Hager, 2001). However, these concepts are based on 2-D steady-state models with imposed force balances. In fact, the dissipative

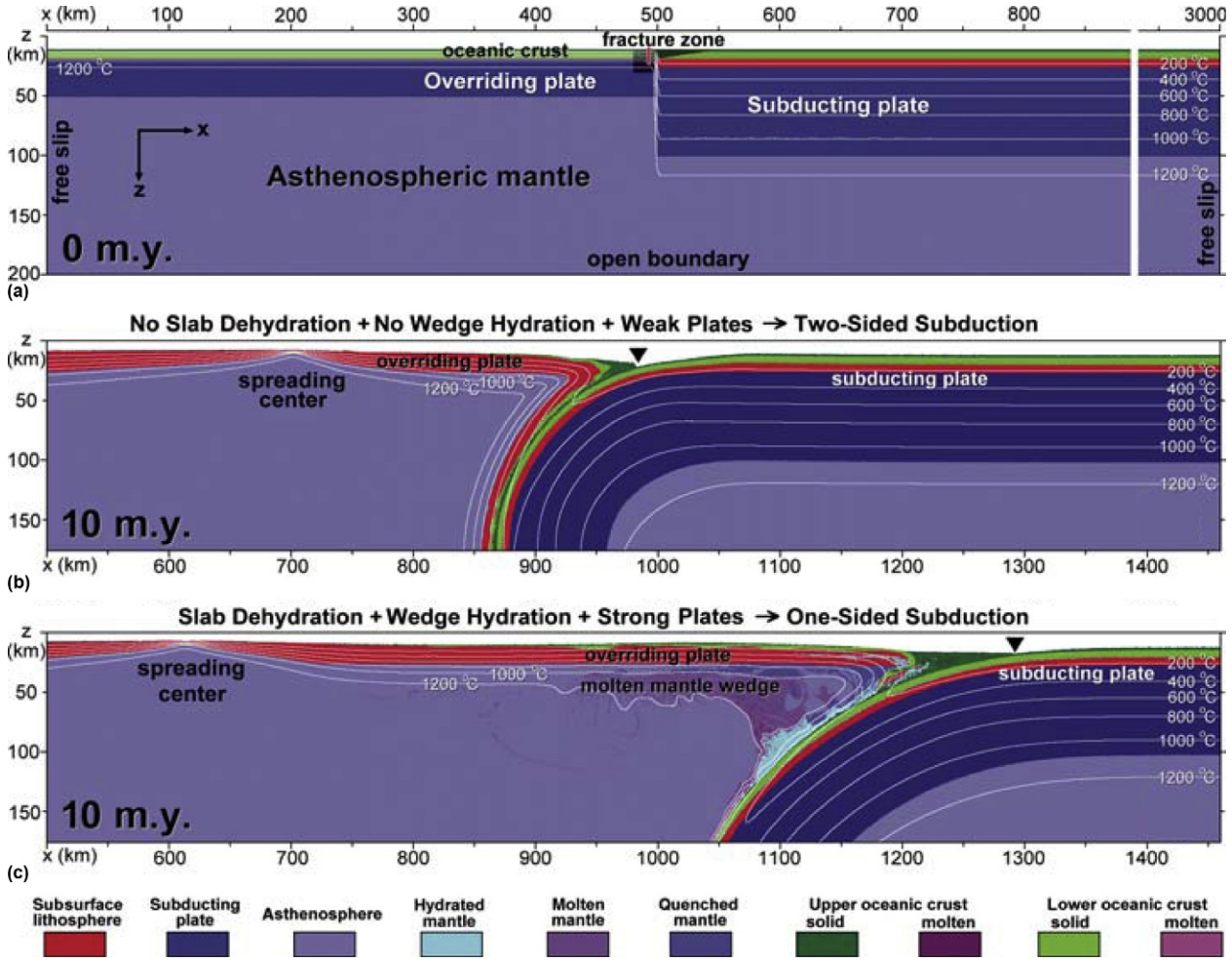


Figure 21 Comparison of subduction flow calculations (model setup, (a)) with imposed weak zone and plastic softening and with (b) and without (c) the effect of water exchange between the slab and the wedge. Although subduction can initiate and proceed with weakening, it typically entrains the overriding plate (b); with slab dehydration, wedge hydration, and melting, the acute corner is lubricated and one-sided subduction is stable (c). Reproduced from Gerya TV, Connolly JA, and Yuen DA (2008) Why is terrestrial subduction one-sided? *Geology* 36(1): 43–46.

effect of such slab bending has recently been inferred to provide only minor resistance; for example, dissipation is minor when accounting for the bending radius of curvature being dependent on plate age and thickness rather than being constant as assumed by Conrad and Hager (2001); moreover, in three dimensions, the extra degree of freedom to evolve the length and curvature of the slab mitigates bending dissipation significantly (Buffett and Becker, 2012; Davies, 2009; Leng and Zhong, 2010). (See also **Chapter 7.09**).

Indeed, this phenomenon of age-independent subduction may just be mysterious from a 2-D and steady-state perspective of convection. A well-developed slab extending across the mantle is certainly capable of drawing a passive ridge into a subduction zone and thus consuming the plate; likewise, trench rollback due to plate foundering would also create 'young' subduction ages. Moreover, the presence of continents and passive margins might trigger subduction at younger than expected ages (Coltice et al., 2012). Thus, the age distribution of subduction zones is possibly a manifestation of time-dependent mantle convection and 3-D plate and crustal evolution. The treatment of temporal variations in plate–mantle systems is therefore a fertile area for future research.

7.07.5.5.4 Melting, crust and depleted lithosphere buoyancy, and the eclogite transition

Although the convective picture of tectonic plates is that they are thickening cold thermal boundary layers, partial melting at mid-ocean ridges induces two chemical effects that counter the plate's negative thermal buoyancy and tendency to sink at subduction zones. In particular, melt extrudes onto the seafloor as basalt, which is less dense than standard mantle peridotites. Moreover, melting extracts incompatible elements from the remaining solid lithosphere, and thus, it dries the lithosphere and depletes it of iron (Hirth and Kohlstedt, 1996), leaving it both stiffer and also chemically buoyant. The net effect is that the lithosphere is harder to subduct than a standard thermal boundary layer; indeed, the lithosphere is frequently not even negatively buoyant when it reaches subduction zones and by all rights should not even subduct. However, basalt reaching greater than about 60 km depth

transforms to eclogite (Ahrens and Schubert, 1975a,b; Davies, 1992; Ito and Kennedy, 1971; Ringwood and Green, 1966), which is denser than lithospheric peridotites and promotes sinking. If basalt plus depleted lithosphere is nonsubductable, then it conceivably piles up over mantle downwellings until the basalt thickness exceeds the eclogite transition or the thermal boundary layer lengthens until its negative thermal buoyancy is sufficient to overcome the positive chemical buoyancy and sink (see also Davies, 1992, 2007a). Either way, once the basalt–eclogite transition is reached, the chemical buoyancy caused by melting is effectively erased and the downwelling current proceeds as a normal cold thermal; that is, in essence, the melting transition and eclogite transition cancel each other out. Therefore, claims that the eclogite transition drives tectonics (e.g., Ringwood, 1972; Ringwood and Green, 1966) tend to ignore the fact that ridge melting hinders it; likewise, claims that melting precludes normal convection neglect that the eclogite transition facilitates it. Taken together, these effects essentially negate each other.

7.07.5.5.5 Trench-parallel flow and rollback

During the 1980s and 1990s, seismic tomography was developed to infer mantle heterogeneity by resolving seismically fast and slow regions that are presumably associated with cold and warm anomalies, respectively (see Volume 1). Seismic tomography thus provides invaluable snapshots, primarily of the mantle temperature field; however, it does not indicate flow direction. In the last decade, measurements of seismic anisotropy have been refined through the analysis, for example, of shear-wave splitting, wherein the differences in velocity between vertical and horizontal polarizations of shear waves are used to infer material anisotropy. Although anisotropy can have many causes, its source in the deeper lithosphere and upper mantle is thought to originate from lattice-preferred orientation in olivine. Such olivine fabric tends to align in the direction of net integrated strain (Kaminski and Ribe, 2001) and thus is indicative of the direction of flow and deformation (see Volume 1 **Chapter 1.09**). One of the more remarkable observations stemming from global shear-wave splitting measurements and upper-mantle anisotropy is that mantle flow directly beneath many slabs at subduction zones is parallel to the trench, not perpendicular to it, as would be expected, given the direction of plate motion (Long and Becker, 2010; Long and Silver, 2008; Russo and Silver, 1994). Trench-parallel flow is perhaps best explained by retrograde slab motion, or trench rollback, whereby the foundering of the slab induces toroidal flow around the slab, that is, the mantle is squeezed sideways around the slab as it falls backward. As even this foundering motion involves a strong downward flow, the dominance of the trench-parallel signal appears to require that near the trench, the mantle is largely decoupled from the downward motion by some lubrication by either a low-viscosity asthenosphere or non-Newtonian behavior (see Long and Becker, 2010). Evidence that slab foundering and rollback are a major component of mantle flow implies that time-dependent plate evolution, such as retrograde motion and plate loss, plays a vital role in the generation and operation of plate tectonics on Earth (e.g., Bercovici and Long, 2014).

7.07.5.6 Divergent Margins: Ridges and Narrow, Passive Upwellings

Basic convection theory predicts that upwellings occur either as columnar plumes rising actively under their own buoyancy or as a very broad background of upwelling ascending passively in response to the downward flux of concentrated cold thermals. Basic convection theory and modeling do not predict focused, shallow, and passive upwellings (i.e., which rise in response to lithospheric spreading motion, rather than by their own buoyancy) analogous to ridges (Lachenbruch, 1976) (see **Chapter 7.08**). That all mid-ocean ridges involve passive upwelling is not necessarily universal (Lithgow-Bertelloni and Silver, 1998). However, the fastest and arguably the most significant ridge, the East Pacific Rise, is almost entirely devoid of a gravity anomaly or deep seismic structure (Forsyth et al., 1998), implying shallow, isostatic support and therefore no deep upwelling current lifting it up (see Davies, 1988a). (Although the interpretation of gravity by itself is nonunique, the lack of a significant free-air gravity anomaly over a topographic feature suggests that the gravity field of the topographic mass excess is being canceled by the field of a nearby mass deficit; this deficit is most readily associated with a buoyant and shallow isostatic root on which the topographic feature is floating.) Moreover, 90% of the structure of the Earth's geoid (the equipotential surface) can be explained by the gravitational attraction of the mass anomalies due only to slabs, implying that other mass anomalies, for example, those due to active buoyant upwellings, are much less significant (Ricard et al., 1993; Richards and Engebretson, 1992; see **Figure 17**). The absence of thermal anomalies beneath ridges is also in agreement with the fact that a simple model of isothermal half-space cooling with $\sqrt{\text{age}}$ satisfactorily predicts seafloor topography under all the oceans (see **Section 7.07.4.2**).

The cause and initiation of narrow ridges are easily as enigmatic as the problems with subduction. The orientation of ridges more or less mirrors the subduction zones that they eventually feed; thus, they may initiate as a strain localization, such as a self-focusing necking instability (Ricard and Froidevaux, 1986), or quite simply an effective tear in the lithosphere. This would suggest that the stresses in the lithosphere due to the pull of slabs are guided considerable distances such that they can be concentrated on particular regions (otherwise, continuous release and diffusion of stresses become manifested as distributed – and thus very unridge-like – deformation). A cold lithosphere, strong by virtue of its temperature-dependent viscosity, makes a plausible stress guide. Strain localization may be due to a variety of nonlinear feedback mechanisms, for example, necking in a non-Newtonian strain-softening medium and weakening due to melting (induced by pressure release melting within the passive upwelling; see Katz, 2008; Spiegelman and McKenzie, 1987; Tackley, 2000d; Turcotte and Morgan, 1992). Additional complexities involving hydrothermal circulation, melt production, and a healing timescale (see also **Section 7.07.6.2**) in non-Newtonian viscoplastic models also predict the formation of transform ridge offsets (Gerya, 2010, 2013).

Finally, while stress guides and strain localization are viable mechanisms for ridge formation, the physics that determines

the distances to which the stresses are guided before causing a ridge – distances that also determine plate size – is not clear and remains a fruitful area of research.

7.07.5.7 Strike-Slip Margins: Generation of Toroidal Motion

While the upwellings, downwellings, divergent and convergent zones in basic convection may or may not appear exactly in the plate-tectonic form of convection seen on the Earth, they at least do exist in both forms of convection. However, while strike-slip motion (i.e., toroidal flow, which also includes plate spin) is very significant in the Earth's plate-tectonic style of convection (see **Sections 7.07.2.1.1** and **7.07.3.7**), it does not exist at all in basic convection with simple rheologies (i.e., with constant or depth-dependent viscosity) and is very weak in convection with slightly more complex rheologies (e.g., with temperature-dependent viscosity; Christensen and Harder, 1991). Thus, the correction to basic convection theory necessary to obtain strike-slip motion is nontrivial; this is because it is not simply a matter of adjusting the convective (poloidal) flow field, but involves generating an additional flow field that would not otherwise exist.

A very basic but important aspect of toroidal motion is that it cannot occur in 2-D models of convection (i.e., models with only x and z coordinates). Since toroidal flow occurs as vortex-type motion in the x–y plane (see **Figure 13**), and since convection is driven by buoyancy forces that point in the z direction, a flow model with convection and toroidal motion needs all three directions; thus, only 3-D convection models can obtain toroidal motion. Two-dimensional models of lithospheric motion in the x–y plane are able to generate toroidal motion (Bercovici, 1993, 1995b, 1998; Bercovici and Ricard, 2005, 2013, 2014) as will be discussed later; however, these do not explicitly involve thermal convection. Another important criterion for platelike toroidal motion is that, apart from plate spin (including global rotation of the lithosphere, which has the largest toroidal power; see **Figure 2**), it is manifested in narrow but intense bands of strike-slip shear akin to a strike-slip margin or transform fault.

7.07.5.7.1 Imposing plates or faults

One way of inducing toroidal motion in mantle flow is to apply plate motion on the surface of an isoviscous mantle; the imposed strike-slip motion at the surface forces toroidal motion in the fluid, or mathematically speaking, it provides a boundary condition to eqn [6] that yields nonnull values of ψ (Hager and O'Connell, 1978, 1979). This method of plate-driven mantle flow has yielded a variety of successful predictions, for example, the observation that the flow lines of the forced motion correlate with slab dip (Hager and O'Connell, 1978). A possibly more self-consistent approach is to drive the plate motion with buoyancy-driven (poloidal) flow; the tractions on the base of a plate are first provided by the poloidal flow, and then, the resulting plate motion subsequently drives toroidal flow. This method can be used to infer the mantle's density and viscosity structure by solving for the internal mass anomalies and viscosity stratification that gives the observed plate motions, as well as the observed geoid and topography (Hager and O'Connell, 1981; Lithgow-Bertelloni and Richards, 1998; Ricard and Froidevaux, 1991; Ricard and Vigny, 1989; Ricard and Wuming, 1991; Ricard et al., 1993; Vigny et al., 1991). It is also possible to use this technique to drive plates and toroidal motion with thermal convection (e.g., Gable et al., 1991; Monnereau and Quéré, 2001; Zhong et al., 2000), and models employing this method give insight into how toroidal and convective/poloidal flows interact. However, methods employing prescribed plate geometries cannot address the problem of how plates themselves are generated and evolve.

A still further step toward self-consistent generation of platelike toroidal motion is to impose faults or narrow weak zones in a model lithosphere, which either overlies a convecting mantle (Zhong and Gurnis, 1995a,b) or moves under its own pressure gradients (Gurnis et al., 2000; Zhong and Gurnis, 1996; Zhong et al., 1998). This approach is largely motivated by the fact that many weak zones persist within the lithosphere for long times and are always available to be reactivated by convective forces; this concept is also addressed later in **Section 7.07.6.1.3**. In these models, platelike motion is best generated by low-friction faults, and the

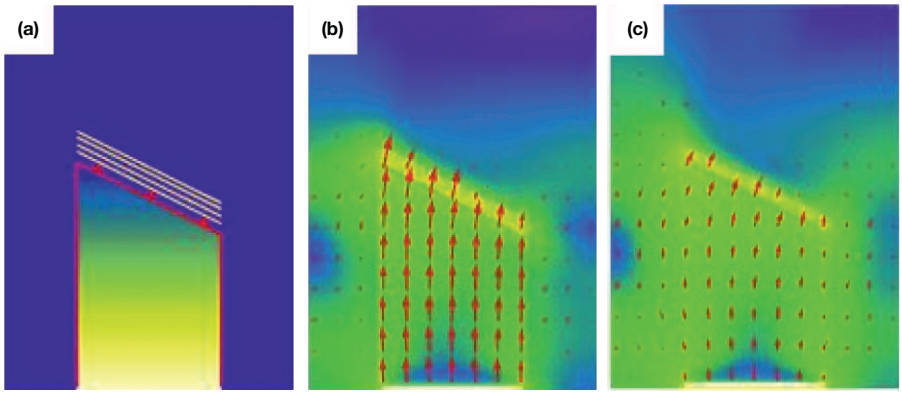


Figure 22 A model of lithospheric flow driven by 'ridge push' (i.e., a lateral pressure gradient due to aging and thickening of the lithosphere) and slab pull with both non-Newtonian rheology and weak faults. (a) The color shading shows lithospheric age (proportional to thickening), slab dip (yellow contours indicate slab depth), and imposed weak faults (red lines). (b) Arrows show velocity field and color shows viscosity for the case with weak faults. (c) Same as (b) but for the case without faults. The weak faults provide lubricated tracks that facilitate much more platelike motion (b) than without the faults (c). See also Zhong et al. (1998). Adapted from Zhong S and Gurnis M (1996) Interaction of weak faults and non-Newtonian rheology produces plate tectonics in a 3d model of mantle flow. *Nature* 383: 245–247.

direction of motion is solely determined by neither the convective (or plate) forces nor the fault orientation; lithospheric motion instead assumes some optimum configuration that takes best advantage of the orientation of both the forces and the faults (Figure 22). Generally, the motion will arrange itself so that strike-slip shear occurs along the low-friction faults (see discussion later in Section 7.07.6.1.2 on the purpose of toroidal flow), and thus, the models generate significant toroidal motion.

However, the methods described earlier permanently prescribe the plate or fault geometries, and thus, these methods do not permit the plates or plate boundaries themselves to evolve from the convective flow. As with the 2-D convection models that strive to generate plateness, 3-D models that attempt to generate toroidal motion (as well as plateness) out of convective flow use a continuum approach with variable viscosity (e.g., Balachandar et al., 1995a; Bercovici, 1993, 1995a,b, 1996, 1998, 2003; Bercovici and Ricard, 2005, 2013, 2014; Cadek et al., 1993; Christensen and Harder, 1991; Foley and Becker, 2009; Kaula, 1980; Ribe, 1992; Stein et al., 2004; Tackley, 1998, 2000a,b,c,d; Trompert and Hansen, 1998; van Heck and Tackley, 2008; Weinstein, 1998; Zhang and Yuen, 1995).

7.07.5.7.2 Why is variable viscosity necessary for toroidal motion?

The reason that the generation of toroidal motion in continuum models requires variable viscosity is somewhat mathematical and thus warrants some theoretical development. As shown earlier (Section 7.07.3.7), if viscosity is constant in viscous buoyancy-driven flows (with homogeneous boundaries), the toroidal motion is zero (see eqn [6]). If viscosity μ is variable, then instead of the equation for toroidal flow being homogeneous as in eqn [6], we have

$$\mu\nabla^2\nabla_{\mathrm{H}}^2\psi = \hat{\mathbf{z}}\cdot\nabla\mu\times\nabla^2\mathbf{v} + \hat{\mathbf{z}}\cdot\nabla\times\left(\nabla\mu\cdot\left(\nabla\mathbf{v}+[\nabla\mathbf{v}]^t\right)\right) \quad [13]$$

(where $[..]^t$ implies transpose of a tensor). Thus, the toroidal field is forced internally (i.e., within the fluid) if $\nabla\mu\neq 0$; if this condition is not met, then the right side of eqn [13] is zero leading to a null solution for ψ as in eqn [6]. This might lead one to believe that toroidal motion can be generated for any variable viscosity, including one that is dependent only on depth (which might occur because of the natural stratification and compression of mantle rock). However, this is not the case. In order to guarantee that the toroidal field obtains some energy from convection, even indirectly, we require that the convective poloidal field ϕ appears on the right side of eqn [13]; that is, the poloidal field gets energy from convective buoyancy (see eqn [5]) and then passes on some of this energy to the toroidal field via viscosity gradients. However, after some algebra, it can be shown that the poloidal field ϕ does not appear at all in eqn [13] if μ is only a function of height z. Therefore, the toroidal field only couples to thermal convection through the poloidal field if $\nabla_{\mathrm{H}}\mu\neq 0$, that is, if viscosity is at least horizontally variable (see also Bercovici, 1993; Hager and O'Connell, 1981; Kaula, 1980; Ribe, 1992; Ricard and Vigny, 1989).

Equation [13] suggests another important condition for generating toroidal motion; in particular, the first term on the right side will be most significant if the horizontal viscosity gradient is orthogonal to the horizontal velocity. A weak zone striking parallel to surface flow will of course satisfy this condition and will thus be a source of toroidal motion and strike-slip shear.

The effect of laterally variable viscosity can be understood qualitatively as well. If convective forces pull or push the upper thermal boundary layer, and part of this boundary layer is weak, then the weak area tends to get moved or deformed more easily than the neighboring stronger zones. The differential motion between the strong and the weak zones can lead to strike slip-type shear and toroidal motion (see Figure 14). However, the mechanism leading to laterally varying viscosity that is sufficient to generate platelike toroidal motion is perhaps the most critical issue in understanding the generation of plate tectonics; this question will be the focus of the rest of the chapter.

7.07.6 Plate Generation Physics

7.07.6.1 Non-Newtonian Mechanics

7.07.6.1.1 What variable viscosities give both plateness and toroidal flow?

As already noted, viscosity is variable by virtue of being possibly dependent on the thermodynamic state, state of stress, grain size, and composition of the medium. Nevertheless, a temperature-dependent viscosity by itself is probably not capable of generating the requisite toroidal flow and plateness. Numerous studies of basic 3-D convection with temperature-dependent viscosity yielded little platelike motion, usually with negligible plateness and toroidal kinetic energies on the order of 10% or less of the total (Christensen and Harder, 1991; Ogawa et al., 1991). The reason these models yielded little toroidal motion is likely due to the upper thermal boundary layer being too cold and stiff to permit any additional motion such as toroidal flow. It may also reflect the fact that since both viscosity and buoyancy are dependent on temperature, the viscosity and poloidal flow fields are in phase, whereas viscosity gradients orthogonal to the poloidal flow are probably optimal for generating toroidal flow (see Section 7.07.5.7.2). Thus, the necessary heterogeneities in viscosity likely arise from other nonlinear rheological mechanisms.

Given the success of creating respectable *plateness* from non-Newtonian power-law or plastic rheologies in 2-D convection models, it stands to reason that similar rheologies would generate significant toroidal motion in 3-D convection models. However, this appears not to be the case. Models using power-law rheologies (even with an excessively large power-law index $n>3$) or plastic rheologies yield modest platelike toroidal power, although the strike-slip shear zones tend to be broad and diffuse (Bercovici, 1993, 1995b; Cadek et al., 1993; Christensen and Harder, 1991; Stein et al., 2004; Tackley, 1998; Trompert and Hansen, 1998; Weinstein, 1998).

It has been proposed that strain softening-type power-law rheologies are not sufficient to excite the necessary shear-localizing behavior and platelike flow and instead that *velocity-weakening* or equivalently *pseudo-stick-slip rheologies* are necessary to generate such motion (Bercovici, 1993, 1995b). A velocity-weakening rheology involves a constitutive law that acts like a normal, highly viscous fluid at low strain rates, but at larger strain rates, the viscosity weakens so rapidly with increasing strain rate that the very resistance to flow itself – that is, the

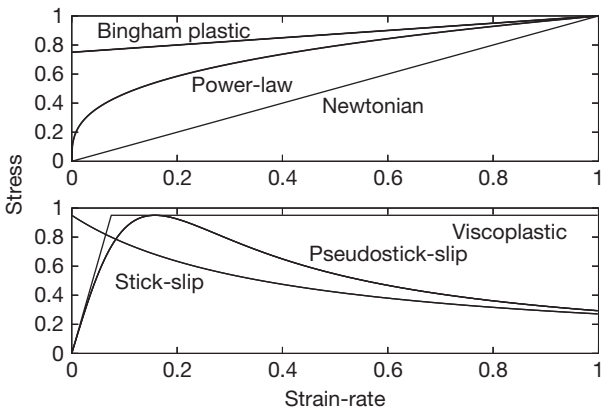


Figure 23 The stress versus strain-rate curves for various rheologies. A Newtonian rheology is represented by a linear constitutive relation between stress and strain rate. A non-Newtonian power-law rheology has a nonlinear constitutive relation in which strain rate goes as stress to some power $n > 1$ (the curve shows the case for $n = 3$ that is typical of deep mantle silicates). Both Bingham plastic and viscoplastic rheologies describe discontinuous stress–strain rate relations whereby, at a particular yield stress, an abrupt transition in flow behavior occurs from strong (immobile or highly viscous) to weak (viscous or with only a maximum strength or allowable stress). Stick-slip and pseudo-stick-slip behaviors make transitions at a peak stress from strong (immobile or highly viscous) to one where strength (i.e., the allowable stress or resistance to deformation) is lost with faster deformation.

stress – decreases; this rheology is most simply represented by the constitutive law

$$\sigma \sim \frac{\dot{e}}{1 + \dot{e}^2/\gamma^2} \qquad [14]$$

where γ is the strain rate at which the stress σ reaches a maximum and begins to decrease with increasing strain rate (**Figure 23**). When employed in various flow models (which permit motion in the x–γ plane and thus toroidal flow), the velocity-weakening rheology yields much more platelike toroidal motion and high plateness than do the power-law rheologies (Bercovici, 1993, 1995b; Tackley, 1998) (see **Figures 24** and **25**).

The velocity-weakening rheology, however, has proved difficult to implement in mantle convection simulations because of its double-valued strain rate (i.e., any given stress has two possible values of strain rate). Convection studies have been, in general, more successful using a simple viscoplastic lithosphere overlying a viscous convecting mantle (Moresi and Solomatov, 1998; Tackley, 2000c; Trompert and Hansen, 1998) including fully 3-D spherical models (Foley and Becker, 2009; Richards et al., 2001; van Heck and Tackley, 2008) (**Figure 26**). The viscoplastic rheology dictates that at low strain rates, material deforms like a very stiff, viscous fluid, but at a given stress (a 'yield stress'), the resistance to motion stops increasing with strain rate and remains constant (**Figure 23**). However, these models produce only moderate platelike behavior (Tackley, 2000c; Trompert and Hansen, 1998), unless used in conjunction with an imposed low-viscosity asthenospheric channel (Richards et al., 2001; Tackley, 2000d) or viscosity reduction due to melting in near-surface mantle upwellings (Tackley, 2000d) (**Figure 27**). Although these viscoplastic models still have difficulty generating isolated strike-slip (toroidal) zones, the inclusion of melting leads to very localized spreading centers with shallow passive upwellings (**Figure 27**). Nevertheless, instantaneous or 'steady-state' rheologies, including the velocity-weakening one, fail to allow inactive or dormant weak zones that can be reactivated, which is possibly a critical feature of plate tectonics; this is discussed further in **Section 7.07.6.1.3**.

7.07.6.1.2 What is the purpose of toroidal motion?

Continuum models that generate toroidal motion from first principles allow us to address an important question: Why does toroidal motion form in the first place, and what role does it play? Since poloidal motion involves upwellings and downwellings, it can enhance its own energy uptake by increasing the rate at which gravitational potential energy is released; this feedback effect is the reason that heated layers go convectively unstable in the first place. However, toroidal flow only dissipates the energy provided: since it involves purely horizontal motion, it does not explicitly enhance gravitational energy release, nor does it, by the same token, transport heat and directly facilitate cooling of the fluid. By all appearances, the toroidal field is superfluous. And yet it is omnipresent in the Earth and is readily generated in fluid models with non-Newtonian (especially velocity-weakening) rheologies.

The reason that toroidal motion is generated appears to relate to the thermodynamic efficiency of surface flow. Examination of the total viscous dissipation in the continuum models of Bercovici (1993, 1995b) shows that even though the toroidal flow field is added to the convectively driven poloidal field, it actually makes the system *less* dissipative than if it were not present (Bercovici, 1995a); that is, generation of toroidal motion *enhances* the efficiency of surface motion. Indeed, the toroidal field establishes itself in focused zones of strike-slip shear that act as lubricated motion guides. The bulk of the fluid outside of these zones of intense toroidal motion remains relatively undeformed and thus entails little dissipation. Within these zones, deformation may be intense, but the viscosity is so low that the dissipation is small. Moreover, the zones of deformation are ideally so narrow and involve so little volume that they make little contribution to the total dissipation generated throughout the whole medium.

7.07.6.1.3 Instantaneous rheologies and plate boundary history

Strongly non-Newtonian rheologies, such as those in the plasticity or velocity-weakening limit, appear to be reasonably successful at generating platelike motion. However, there are a few fundamental problems with this approach. First, these very nonlinear rheologies tend to be ad hoc. For example, while the plasticity formalism is based on known plastic yielding of rocks, it requires a much lower low yield stress or friction coefficient relative to the actual strength of rocks. Moreover, the velocity-weakening mechanism is an assumed law since there is no steady-state creep law that displays this behavior; it is if anything a steady-state representation of a more complex time-dependent self-softening mechanisms (see **Section 7.07.6.2**). Indeed, laboratory experiments on rock deformation at moderate temperatures show much more complex behavior than simple instantaneous rheologies (e.g., time

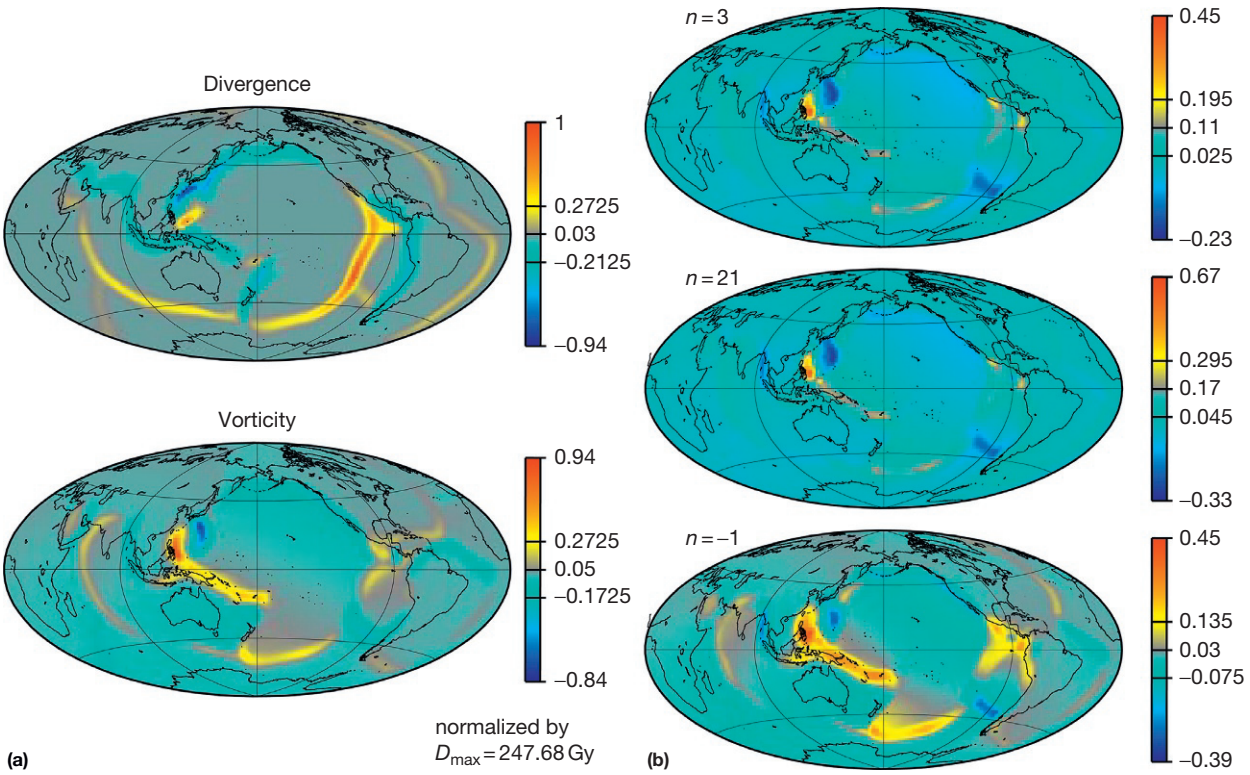


Figure 24 Divergence rate and vertical vorticity (rate of spin and strike-slip shear) of present-day plate motions (a), representing poloidal and toroidal motion, respectively. Red and yellow divergence is positive over spreading centers and blue is negative, or convergent, over subduction zones. Red and yellow vorticity represents left-lateral strike-slip motion and counterclockwise spin, while blue is right-lateral motion and clockwise spin. Fluid-dynamic calculations use the Earth's divergence as a source–sink field to drive flow in various non-Newtonian shallow-layer model lithospheres, thereby generating toroidal motion that is shown in terms of vorticity (b). Power-law rheologies with indexes $n = 3$ and even as high as $n = 21$ (where stress goes as strain rate to the $1/n$ power; see eqn [11]) generate vorticity fields that do not compare well with the Earth's vorticity field. A self-lubricating or pseudo-stick-slip rheology, represented by $n = -1$ (see eqn [14]), however, generates vorticity that compares well with the Earth's present-day case. Adapted from Bercovici D (1995) A source–sink model of the generation of plate tectonics from non-Newtonian mantle flow. *Journal of Geophysical Research* 100: 2013–2030.

evolution of grain size and texture, as well as interaction between different deformation mechanisms; see Karato, 2008).

Second, these non-Newtonian rheologies involve an *instantaneous* response because viscosity responds only on the instantaneous strain-rate field. Thus, weak zones, as proxies for plate boundaries, only exist as long as they are being deformed; once deformation ceases, the boundaries instantaneously vanish (Gurnis et al., 2000; Zhong et al., 1998). However, it is known that while new plate boundaries continue to be formed, old inactive ones persist as fabric in the lithosphere and even provide intrinsic weak zones that are preferred sites at which to reactivate deformation, especially for the initiation of new subduction zones in cold strong lithosphere (Gurnis et al., 2000; Hall et al., 2003; Lebrun et al., 2003; Toth and Gurnis, 1998). Thus, plate boundaries have memory: they can remain weak even without being deformed and be advected with the material in which they are embedded. This implies that the weakness is an actual transportable property, like temperature, chemical concentration, or grain size. Moreover, such weak zones can persist for extremely long times, possibly much longer than the typical convective timescale of 10^8 years (a typical convective overturn time) (Gurnis et al., 2000); this suggests that the property that determines plate boundary weakness obeys a much longer timescale than does convection.

7.07.6.2 Dynamic Weakening and Damage Physics

That plate boundaries are caused by some weakening property rather than an instantaneous rheological response correlates with the origins of the velocity-weakening and localizing feedback. There are various candidate feedback mechanisms that we discuss next.

7.07.6.2.1 Thermal self-weakening

The velocity-weakening rheological law expressed in eqn [14] is a simplified, 1-D, steady-state representation (Bercovici, 1993; Whitehead and Gans, 1974) of the classical self-softening feedback arising from the coupling of viscous heating with temperature-dependent viscosity (Schubert and Turcotte, 1972). With this feedback mechanism, deformation causes frictional heating, which warms and weakens the material; the weak zone is more readily deformed causing a concentration of deformation and therefore more heating, weakening, etc. In this case, a weak zone corresponds to a temperature anomaly, and thus, temperature acts as the transportable time-

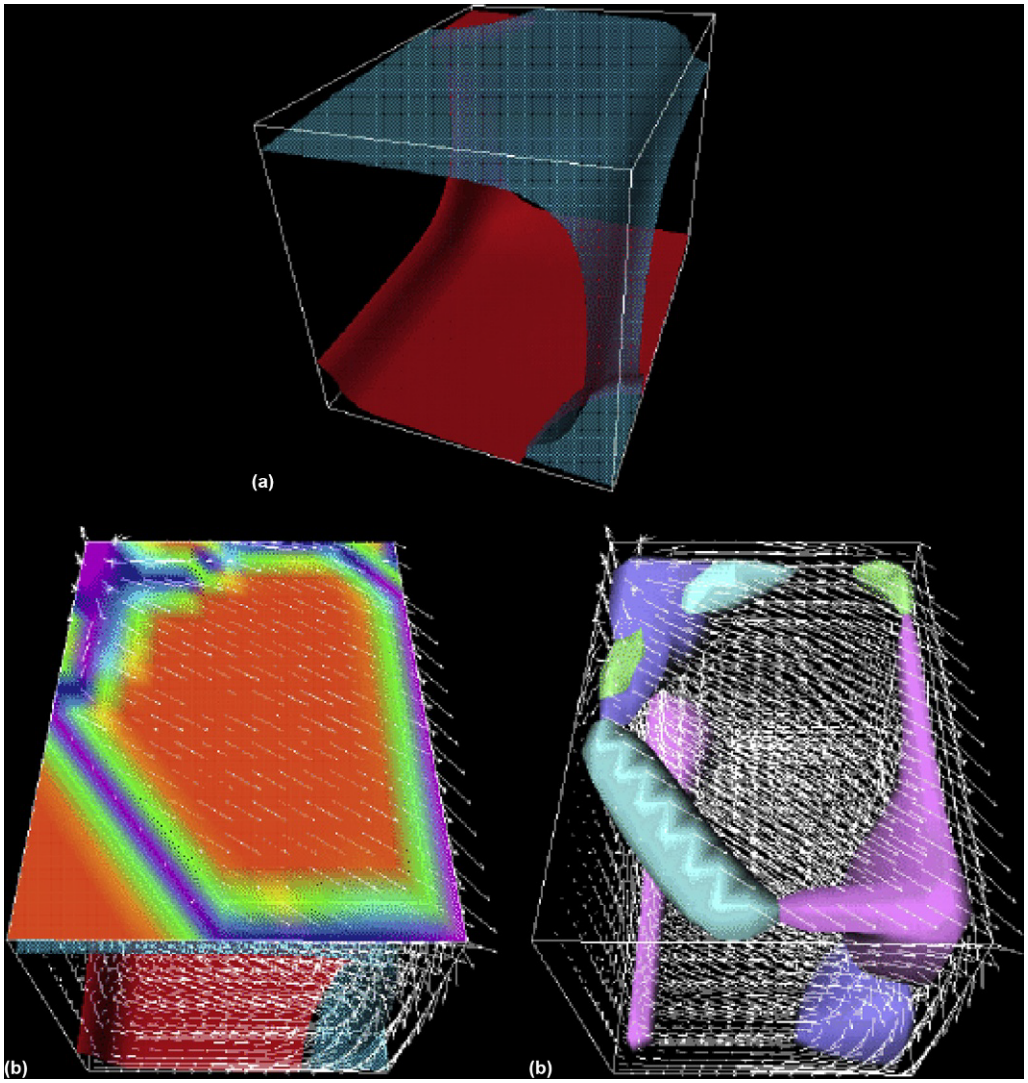


Figure 25 Generation of platelike motion in lithospheric flow driven from below by a convective-type buoyancy field. (a) The temperature field from simple basally heated convection. This is used as a static source of buoyancy to drive flow that then drags an overlying layer of fluid with stick-slip rheology (eqn [14]). The sharp changes in viscosity (b) surrounding nearly isoviscous (red) blocks, along with nearly uniform velocity fields in these blocks, are suggestive of highly platelike flow. (c) Isosurfaces of horizontal divergence (poloidal flow) in the lithospheric layer (pink and purple) and vertical vorticity or toroidal flow (light blue and green). The elongated blue surface corresponds to a long, fault-like concentration of strike-slip shear. Adapted from Tackley P (1998) Self-consistent generation of tectonic plates in three-dimensional mantle convection. *Earth and Planetary Science Letters* 157: 9–22.

dependent property whose history can be retained for a finite period of time.

?A3B2 tlsb=0.01w?>However, when this viscous heating-based mechanism is incorporated into flow models, it is not overly successful at generating large-scale platelike motions. In lithospheric flow models, this mechanism gives relatively weak localization and diffuse toroidal motion (Bercovici, 1998) (see **Figure 28**). In convection models, viscous heating can be significant on small scales in the lithosphere and, with temperature-dependent viscosity, can lead to modest local concentrations of toroidal motion (Balachandar et al., 1995a,b; Zhang and Yuen, 1995); however, the surface motions and strength distributions are not platelike, and the global toroidal energy is typically very small. This lack of platelike motion probably occurs for several reasons. First, the temperature anomalies necessary to reach a velocity-weakening regime are excessively large, near the melting temperature, which is well beyond that attainable with viscous heating in these systems (Bercovici and Karato, 2003); that is, dissipative thermal anomalies are self-limiting because dissipative heating is proportional to the local viscosity, which drops precipitously with increasing temperature (Fleitout and Froidevaux, 1980; Yuen et al., 1978). Secondly, and for similar reasons, the temperature anomalies due to viscous dissipation are generally much smaller than temperature anomalies due to heat that powers convection (e.g., radiogenic heating). Thus, the temperature drop across the cold upper thermal boundary layer is usually much larger than any temperature anomaly due to viscous heating, causing this boundary layer to continue to be stiff and immobile.

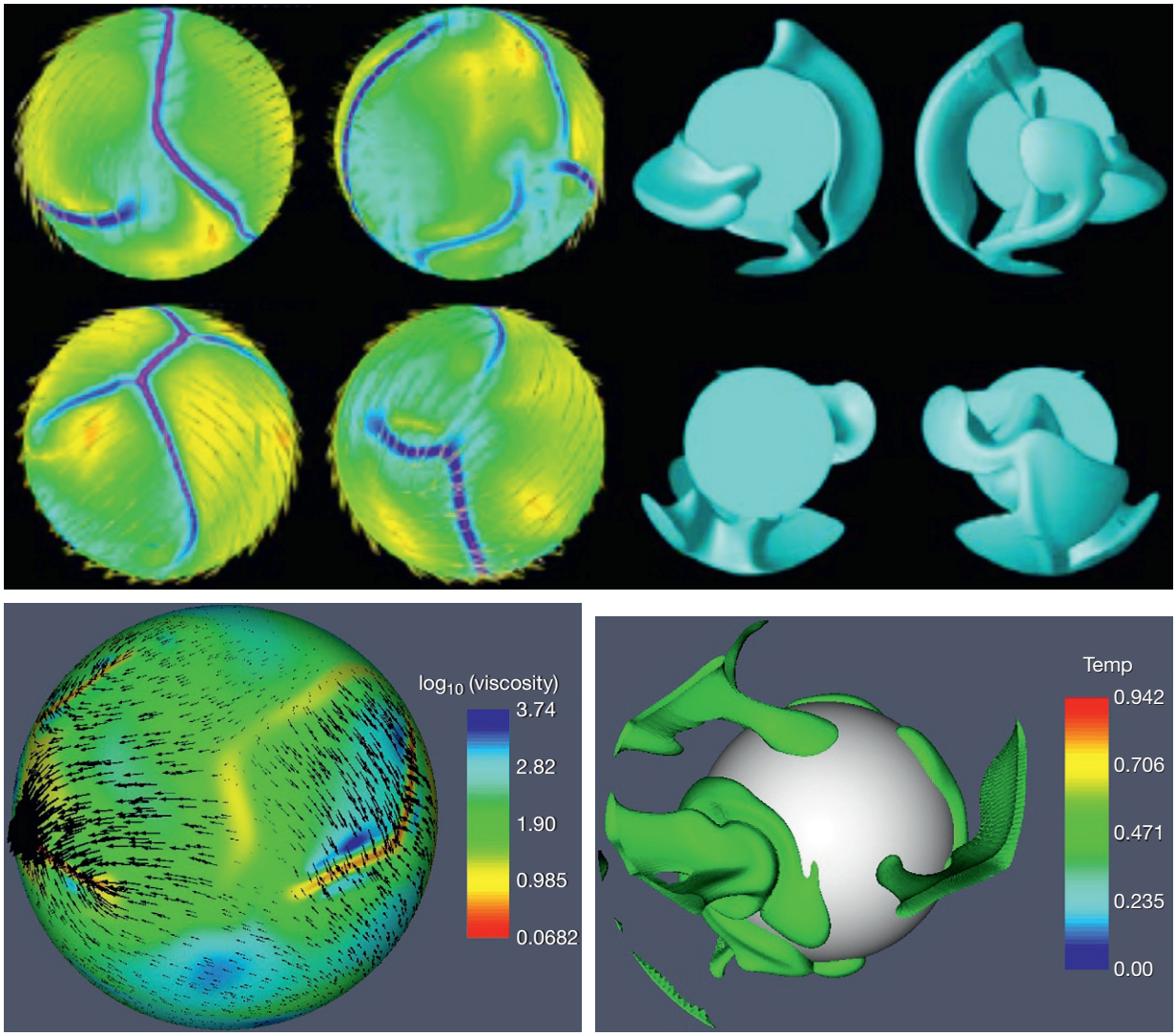


Figure 26 Three-dimensional spherical-shell convection with a plastic-type lithospheric rheology from two different studies of van Heck and Tackley (2008) (top) and Foley and Becker (2009) (bottom) showing platelike behavior. Left panels show surface viscosity with velocity vectors superimposed. Right panels show isothermal surfaces and in particular cold downwellings. Note the passive divergent and rheological weak zone forming midway between the two major downwelling regions.

Moreover, if the Earth's weak zones were due to temperature anomalies (and were fortuitously not overwhelmed by convective temperature anomalies), they would diffuse away on the order of a few million years (assuming plate boundaries are of order 10 km wide and thermal diffusivity is $10^{-6}\,m^2 s^{-1}$), which is not observed (Gurnis et al., 2000). Finally, if plate boundaries, in particular strike-slip margins, were caused by viscous heating, then they should be perceptible with heat flow measurements; however, this is generally thought not to be the case (Lachenbruch and Sass, 1980) (cf. Thatcher and England, 1998).

Nevertheless, while viscous heating alone is unlikely to provide a sufficient sustained localizing feedback mechanism, it could play a role in assisting localization associated with grain-size reduction (Kameyama et al., 1997) (see also **Section 7.07.6.2.3**) and in helping initiate subduction (Thielmann and Kaus, 2012) (see also **Section 7.07.5.5.1**).

7.07.6.2.2 Water and void weakening

An appealing yet speculative plate generation model invokes water as the weakening agent; this is motivated by the oft-stated hypothesis that plate tectonics only exists because water lubricates plate boundaries, which also partially explains the lack of plate tectonics on the terrestrial planets devoid of water (Lenardic and Kaula, 1994, 1996; Tozer, 1985). In particular, water is assumed to provide pore pressure reduction of friction and/or lubrication of slip zones by, for example, serpentinization (Escartin et al., 2001). Moreover, if weak zones are due to water, then their longevity would be controlled by chemical diffusivity of hydrogen in minerals, which is faster than most chemical diffusivities but much slower than thermal diffusivity (Brady, 1995), in which case weak zones could survive over geologic timescales.

Water is certainly ingested by subducting slabs and plays a critical role in maintaining sufficient weakness at subduction

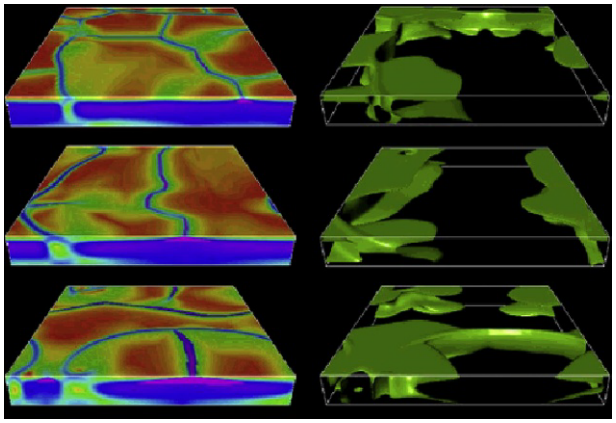


Figure 27 Three-dimensional simulations of convection with a viscoplastic rheology and including viscosity reduction by melting. Convection is driven entirely by internal heating, and thus, cold downwellings are the dominant current. The left panels show the viscosity fields, where purple and blue are low viscosity and red and yellow are high viscosity. The right panels show the associated cold isothermal surfaces that outline the downwelling convecting currents that drive motion in the system; therefore, any low-viscosity surface region coinciding with a cold isotherm is a convergent zone. The platelike geometry is apparent in the surface viscosity fields (left panels); in particular, the focused spreading centers overlying melt lenses (purple zones) are diverging passively, that is, they are not associated with any active upwelling rising from deep in the layer. However, strike-slip motion is relatively weak and not isolated, being associated primarily with obliquity and offsets in the spreading centers. Adapted from Tackley P (2000) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 2. Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems* 1: 1026. http://dx.doi.org/10.1029/2000GC000043.

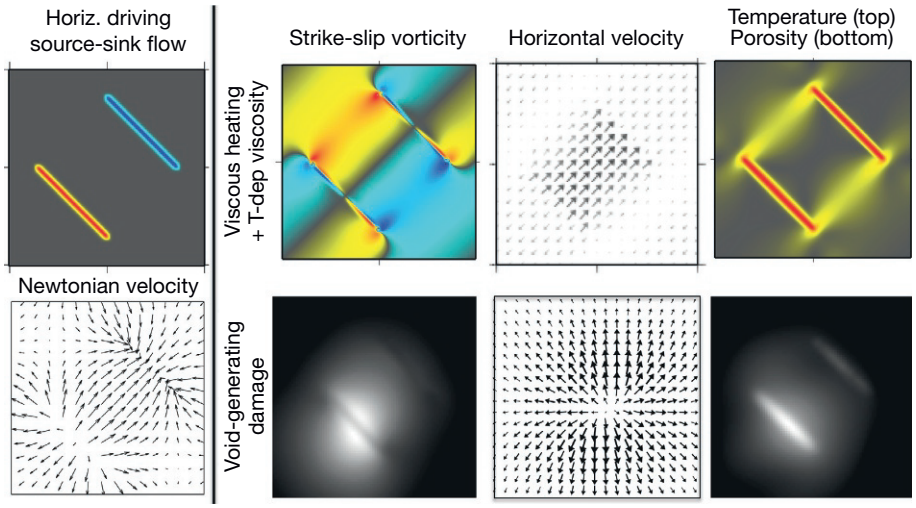


Figure 28 Summary of two separate source–sink model studies of shallow lithospheric flow with different localization mechanisms. The driving source–sink field equals the horizontal velocity divergence and has amplitude of ± 1; the velocity field corresponding to pure source–sink flow, as occurs with a Newtonian uniform rheology, is shown for comparison. The top row of calculated fields (right of the vertical line) is for the case with viscous heating and temperature-dependent viscosity; the generated toroidal vorticity field is relatively weak with an amplitude of ± 0.2 (i.e., only 20% of the driving source–sink field), leading to modest platelike flow yet still diffuse strike-slip shear in the corresponding velocity field and weak temperature and viscosity anomalies over the strike-slip margins. The bottom row of calculated fields is for the case of void-generating damage, as a model for opening up cracks and pore space with which to ingest a weak phase such as water; in this case, dilational flow augments poloidal motion over the driving source and leads to very unplatelike monopolar flow, with poor platelike structure in both the vorticity field and the porosity field (which controls the effective viscosity). Adapted from Bercovici D (1998) Generation of plate tectonics from lithosphere-mantle flow and void-volatile self-lubrication. *Earth and Planetary Science Letters* 154: 139–151; Bercovici D and Ricard Y (2005) Tectonic plate generation and two-phase damage: Void growth versus grainsize reduction. *Journal of Geophysical Research* 110: B03401. http://dx.doi.org/10.1029/2004JB003181.

zones, in addition to the asymmetry of slab downwelling (Crameri et al., 2012; Faccenda et al., 2009, 2012; Gerya et al., 2008; Ranero et al., 2003; see also **Section 7.07.5.5.2**). However, applying the water weakening mechanism to understand shear localization at all plate boundaries is fraught with various limitations. First, frictional reduction by pore pressure is relevant for brittle failure and frictional sliding, which occurs only over shallow depths that are already the weakest portions of the lithosphere (**Figure 19**). The lithosphere reaches its peak strength well below the top roughly 10 km and undergoes complex transitions to semiductile and ductile behavior, where frictional sliding plays little to no role. Second, the

availability of water to weaken the deep strong lithosphere is problematic. Given melting at ridges and extraction of water into the melt phase, most of the lithosphere is probably as dry as on nonplate planets such as Venus (Hirth and Kohlstedt, 1996). Thus, if water is to play a weakening role, it must be ingested from the surface to great depths, down to 100 km or so. The great pressures at these depths cause a very large lithostatic pressure gradient against which fluid has to be pushed, in addition to closing down fluid pathways to negligible permeabilities; given that tectonic stresses are orders of magnitude smaller than lithostatic pressures, it is unlikely they can induce deep pressure lows and hold open fluid pathways against the weight of the lithosphere. One possible solution to this quandary is deep thermal cracking at fracture zones emanating from mid-ocean ridges, which can possibly allow water to be ingested to moderate depths and serpentinize lithosphere in these zones (Korenaga, 2007). However, thermal stresses at best allow water to be ingested to moderate depths (a few tens of kilometers), leaving much of the lower strong lithosphere, and lithosphere far from fracture zones, unaffected. Moreover, the ingestion of water and serpentinization occur only before thermal stresses relax away moderately rapidly, and thus, there is no feedback by which weakening focuses deformation, which causes more weakening, thus focusing deformation further, etc.

Indeed, in order for water to induce a shear-localizing feedback, it must not only weaken the material into which it is ingested, but the ingestion and water content must, in turn, be enhanced by the deformation itself. In particular, ingestion of water may be controlled by rate of deformation when voids are created by damaging of the material, such that these voids take up water, which causes weakening, which in turn focuses deformation, more damage, etc. The physics of this process was elaborated on by 'two-phase damage' theory (Bercovici and Ricard, 2003; Bercovici et al., 2001a,b; Ricard and Bercovici, 2003; Ricard et al., 2001) wherein the energy necessary to create a void or a microcrack (Griffith, 1921) is treated as the surface energy on the interface between a host phase (rock) and the void-filling phase (e.g., water). Deformational work is used to create this interfacial energy by generating more voids, thus inducing weak zones on which deformation localizes. This mechanism leads to a spectrum of shear-localizing behavior, from diffuse to very sharp shear zones (Bercovici and Ricard, 2003; Bercovici et al., 2001b), and also predicts failure in low-cohesion rocks that undergo dilational or compactive/cataclastic damage (Ricard and Bercovici, 2003). However, when this mechanism is placed at high lithostatic pressures typical of the mid-to-deep lithosphere, it is suppressed relative to other deformation mechanisms, given the difficulty of opening new pores (Landuyt and Bercovici, 2009a). Moreover, when this mechanism is employed in 2-D lithospheric flow models, wherein imposed source–sink poloidal flow is used to generate strike-slip toroidal motion, it does not yield platelike solutions (Bercovici and Ricard, 2005); in particular, the dilational motion from void generation augments poloidal flow at the expense of toroidal motion, thereby leading to very unplatelike flow (see **Figure 28**).

In the end, water weakening remains problematic for generating weak zones that extend across the entire lithosphere and for generating toroidal motion. Certainly, water-enhanced melting is likely critical for melt localization at ridges and weakening of the mantle wedges at subduction zones (see **Sections 7.07.5.5.2** and **7.07.5.6**). Whether water directly leads to the formation of all plate boundaries especially the most enigmatic 'toroidal' ones is, however, doubtful, although its indirect effect on surface conditions remains a possibility (see **Section 7.07.7.3**).

7.07.6.2.3 Grain-size weakening

Observations of localized shear in mantle peridotitic mylonites – where large strain correlates to highly reduced mineral grain sizes – at all types of plate boundaries (e.g., Dijkstra et al., 2004; Etheridge and Wilkie, 1979; Furusho and Kanagawa, 1999; Ildefonse et al., 2007; Jin et al., 1998; Skemer et al., 2010; Warren and Hirth, 2006; White et al., 1980) have prompted much activity in exploring grain size-dependent shear-localizing feedback mechanisms (e.g., Braun et al., 1999; Kameyama et al., 1997; Montési and Hirth, 2003). In this case, a self-weakening positive feedback occurs because of the interaction of grain size-dependent rheologies (such as diffusion creep or grain-boundary sliding; see Hirth and Kohlstedt 2003) and grain-size reduction driven by deformation through dynamic recrystallization (e.g., Derby and Ashby, 1987; Doherty et al., 1997; Karato et al., 1980; Lee et al., 2002; Shimizu, 1998; Urai et al., 1986). However, this localizing feedback mechanism is paradoxical because grain-size reduction by recrystallization is generally thought to occur only in dislocation creep, which is independent of grain size, while rheological softening by grain-size reduction only occurs in diffusion creep, when the grains cannot be reduced (De Bresser et al., 1998, 2001; Etheridge and Wilkie, 1979; Karato and Wu, 1993; see **Figure 29**). Large grains undergoing dislocation creep can shrink through dynamic recrystallization, while small grains undergoing diffusion creep coarsen by normal grain growth; grain sizes thus tend to evolve toward the boundary between diffusion and dislocation creep, and the rheology therefore tends to become anchored closed to the boundary, on the piezometric curve. Near this boundary, complex interactions can ensue, leading to effective rheologies that are dependent on both grain size and stress. Such interactions can be due either to mixing of creep mechanisms over grain-size distributions that span the boundary (e.g., Rozel et al., 2011) or to unique mechanisms like grain-boundary sliding (e.g., see Hansen et al., 2012; Hirth and Kohlstedt, 2003); however, these effects are not necessarily distinguishable in that both mixing and grain-boundary sliding lead to similar grain-size and stress dependence (see Hansen et al., 2012; Rozel et al., 2011). Nevertheless, these interactions are bound close to the diffusion–dislocation boundary where grain-size reduction and thus localization feedbacks are restricted (Rozel et al., 2011).

In addition to problems with the grain softening feedback mechanism, grain growth and healing of weak zones in single-mineral or single-phase systems are quite fast (Karato, 1989), which would cause fine-grained weak zones in the lower lithosphere to vanish in less than a million years (see Bercovici and Ricard, 2012). Thus, grain-size reduction in these simple systems cannot by itself support long-lived plate boundaries.

However, many of the problems with this mechanism stem from its basis in monomineralic single-phase systems. That lithospheric and crustal rocks are polymineralic multiphase systems like peridotite (with at least both olivine and pyroxenes) may play an important role in localization and plate

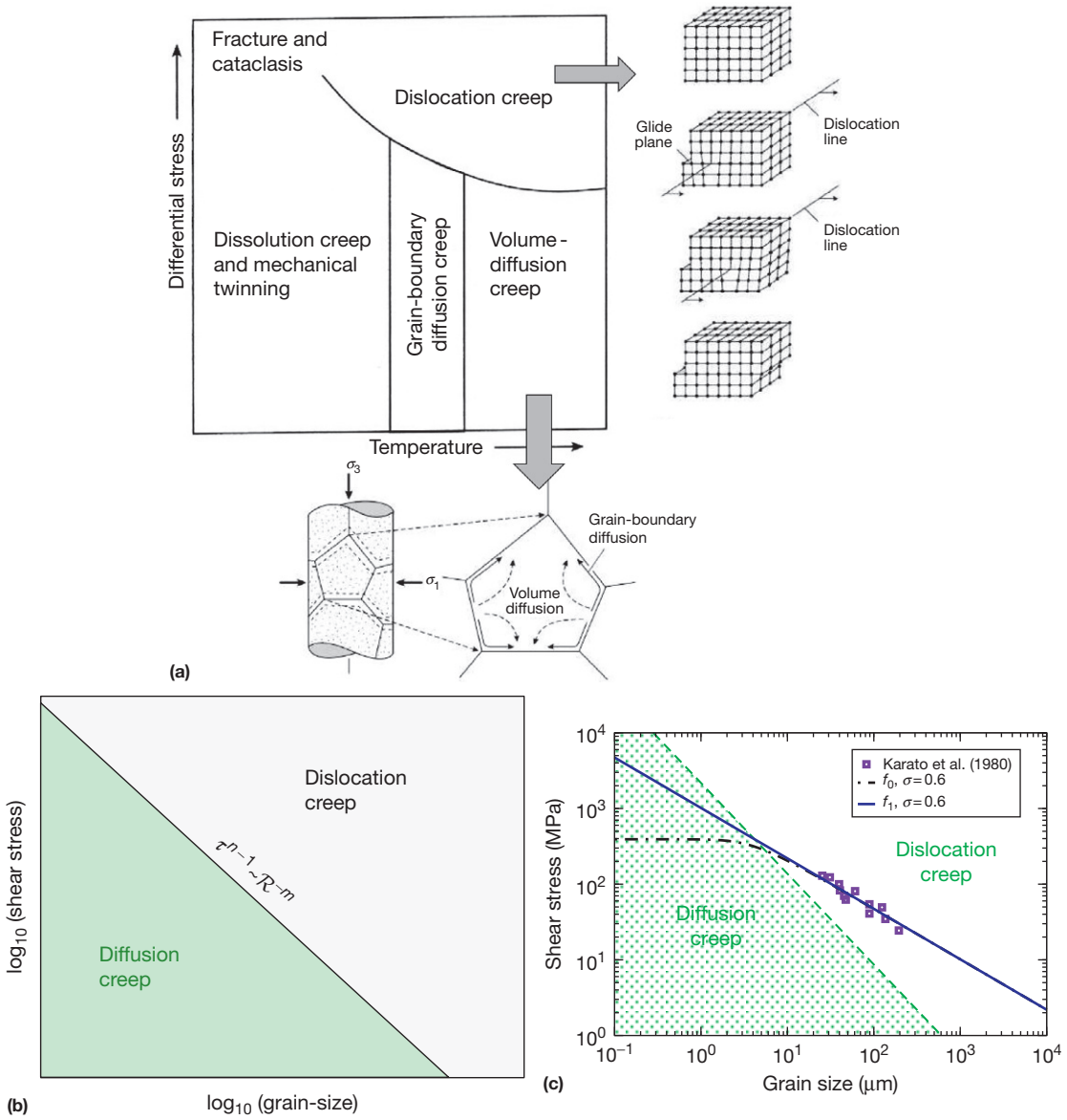


Figure 29 (a) Deformation map of stress–temperature space showing different creep and failure mechanisms; (b) stress–grain size space sketch to show typical creep mechanism transition as a function of grain size (the transition occurs where the strain rates of each mechanism are equal and thus, $\dot{e} \sim \tau^n \sim \tau/\mathcal{R}^m$ and thus, the boundary is given by $\tau^{n-1} \sim \mathcal{R}^{-m}$); and (c) a stress–grain size deformation map and steady-state rheology predicted by the single-phase grain-damage theory of Rozel et al. (2011), where the solid blue curve is for the case where damage only exists in dislocation creep (as expected) and the black dashed-dot curve is if damage can somehow persist into the diffusion creep regime. The two-phase Zener pinning damage model (Bercovici and Ricard, 2012) provides a mechanism for damage to persist well into the diffusion creep regime.

boundary formation. Indeed, many observations of peridotitic mylonites show evidence of secondary-phase pinning in which reduced pyroxene grains appear to hold the olivine in permanent diffusion creep (Herwegh et al., 2011; Linckens et al., 2011; Mehl and Hirth, 2008; Olgaard, 1990; Warren and Hirth, 2006). In this regard, Bercovici and Ricard (2012) recently proposed a generalized two-phase grain evolution and damage theory, which provides a new mechanism that potentially solves the grain-weakening paradox. This theory treats the coupling between damage and Zener pinning in two-phase rock mixtures like peridotite. In essence, pinning in mixtures arises because immiscible phases block each others' grain-boundary migration and hence stunt grain growth. In two-phase theory, the pinning or blocking surface is represented by the interface between phases. This interface undergoes damage by stretching or rending, which reduces the size of the pinning surfaces, which in turn pins grains to smaller sizes regardless of creep mechanism. Damage thus indirectly reduces grain size even in diffusion creep, wherein the material gets weaker as grains shrink, thereby allowing a very effective shear-localizing feedback consistent with observations of mylonites.

This two-phase grain-damage and pinning model has also been tested in 2-D lithospheric source–sink flow models

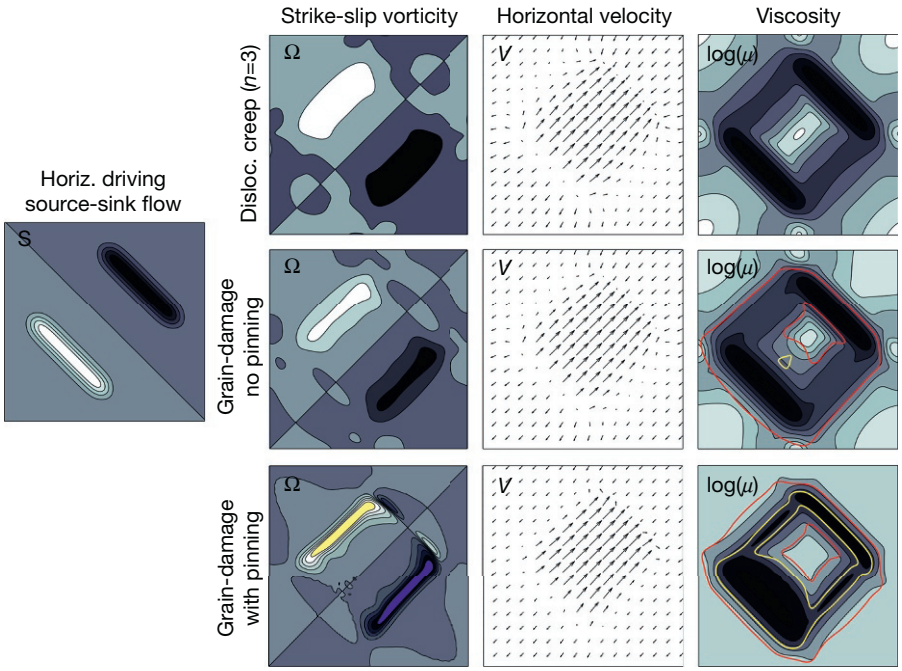


Figure 30 Source–sink-driven flow with simple dislocation-creep power-law rheology ($n=3$), grain damage without pinning, and grain damage with pinning. Yellow and indigo on the Ω contours show saturation above ± 1. Yellow contours on μ plots bound where the medium is more than 99% in diffusion creep (red is the 75% bound). Adapted from Bercovici D and Ricard Y (2013) Generation of plate tectonics with two-phase grain-damage and pinning: Source–sink model and toroidal flow. *Earth and Planetary Science Letters* 365: 275–288.

(Bercovici and Ricard, 2013). Grain-size reduction and damage without pinning (as in Rozel et al., 2011) lead to very weak localization and toroidal motion, similar to that for basic dislocation-creep power-law rheologies (**Figure 30**, top rows). However, the inclusion of pinning and interface damage allows for intense localization and toroidal motion and highly platelike behavior (**Figure 30**, bottom row).

The pinning effect also promotes the longevity of inactive weak zones; in particular, after deformation ceases, grain growth and healing are essentially stopped or retarded by the small pinning bodies, which in themselves grow extremely slowly (especially if moderately dispersed) (Bercovici and Ricard, 2012). Moreover, the damage model shows that the inheritance of weak zones can have a major influence on plate and plate boundary evolution after plate motion changes in 2-D flow models (Bercovici and Ricard, 2013, 2014); this effect provides a simple theory for both the emergence of plate tectonics in the Archean and modern and even abrupt plate reorganizations (Bercovici and Ricard, 2014; Bercovici et al., 2015) (see **Figures 31** and **32**). In the end, the presence of pinning in polycrystalline or two-phase materials permits damage and softening to coexist in a positive shear-localizing feedback while also allowing for long-lived dormant weak zones.

7.07.6.2.4 Summary thoughts on weakening and damage

It is unlikely that any one shear-localizing mechanism will be the so-called magic bullet of plate boundary formation and plate generation. A distinct possibility may be that a combination of rheological mechanisms will be necessary. While brittle failure at shallow depths does not dictate the major weakening effect (since it is occurring in the weakest region of the lithosphere), shallow and intermediate-depth cracks sustained by water might help nucleate self-weakening zones at greater depth. The pervasiveness of mylonitic shear zones, however, is evidence that grain size-dependent damage mechanisms are acting across the deeper strongest part of the lithosphere, and these are likely the rate-limiting processes to generate plate boundaries. Grain damage with pinning in polycrystalline or two-phase mixtures appears to be a promising mechanism for such localization and boundary formation, although it requires further testing in global geodynamic models.

In the end, the simplest ingredients for plate generation might be damage and grain-size reduction (augmented by pinning, as mentioned earlier) in addition to melt-induced weakening at divergent and convergent zones. Although the effect of water on melt weakening is manifested by its reduction of melting temperature, its influence on grain damage is less obvious. One possibility is that water on Earth's surface promotes the carbon cycle and a cool surface, which thus impedes healing and advances the influence of damage and weakening; this hypothesis is central to one explanation for the dichotomy between Earth and Venus (Landuyt and Bercovici, 2009b) and for predicting the occurrence of plate tectonics on super-Earths (Foley et al., 2012), as will be discussed later (see **Section 7.07.7.3**).

7.07.7 Plate Generation on Earth and Other Planets

7.07.7.1 Origin of Plate Tectonics and Archean Tectonics

As discussed in **Section 7.07.2.2**, a critical moment in Earth's evolution is the initiation of plate tectonics itself. Although evidence for the first appearance of plate tectonics is sparse

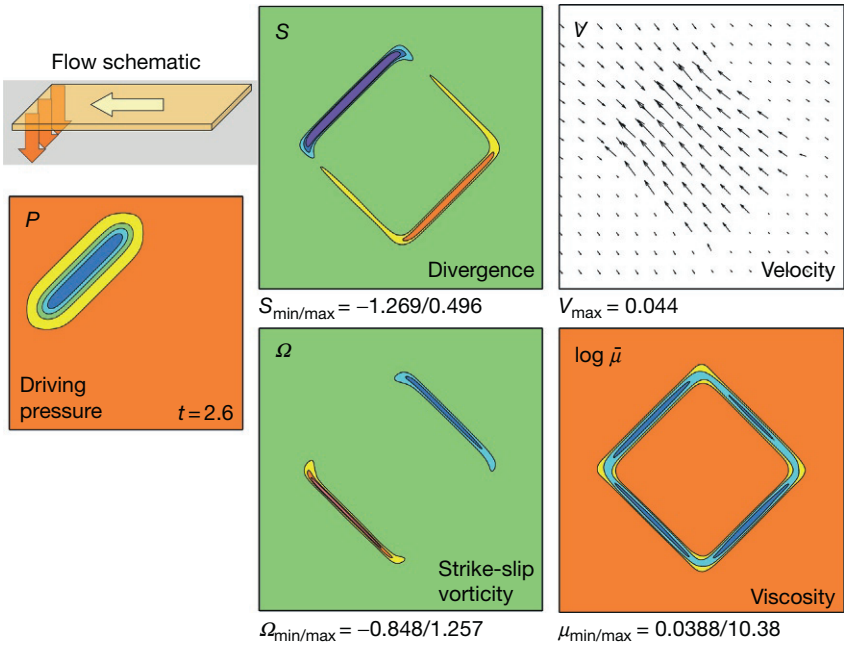


Figure 31 A calculation of pressure-driven flow in a thin 2-D horizontal-layer model of the lithosphere, where the pressure low is akin to a single subduction zone (see schematic frame). The rheology of the layer is governed by the grain evolution, damage, and pinning model of Bercovici and Ricard (2012, 2013), which is also shown in **Figure 30**. In this case, the low-pressure zone P is imposed and then rotated about a vertical axis by 90° three times (with roughly 10 My dimensional time between rotations to develop damaged weak zones) as an idealization of intermittent and chaotic subduction during the early Archean. The bands of damage induced by the pressure low from a previous orientation are long-lived, inherited, and amplified by the lithospheric flow of the next orientation, resulting in localized but passive bands of strike-slip vorticity Ω and positive divergence S (red and yellow contours; convergence with $S < 0$, indicated by blue contours, is actively driven by the pressure low). Thus, a complete plate arises with a contiguous weak plate boundary (indicated by viscosity $\bar{\mu}$) while only being driven by subduction, which is the final state shown (with final dimensionless time indicated on the P frame). Dimensionless extrema for contours or vectors are indicated below each frame, save P, which is always between 0 and 1. This model is proposed to explain the emergence of plate tectonics in the Archean, from protosubduction 4Ga to global tectonics by ~3 Ga. Adapted from Bercovici D and Ricard Y (2014) Plate tectonics, damage and inheritance. *Nature* 506: 513–516. http://dx.doi.org/10.1038/nature13072.

and indirect, it appears that protosubduction existed roughly 4Ga, but global tectonics possibly did not become pervasive for a billion years later (see **Section 7.07.2.2**). Plate generation models have recently begun to explore this question of the onset of subduction and plate tectonics, and it remains a fruitful area of future research (e.g., Bercovici and Ricard, 2014; Korenaga, 2006, 2013; Moyen and van Hunen, 2012; Sizova et al., 2010; van Hunen and Moyen, 2012; van Hunen and van den Berg, 2008) (see also **Figure 31**).

The nature and evolution of the plate–mantle system during the Archean also remain enigmatic. Various lines of evidence imply that plate velocities and crustal production were slower in the Archean (see Korenaga, 2006). Thus, plate motions may have been impeded by greater lithospheric stability and strength through melting and production of buoyant crust and depleted lithosphere (Davies, 1992; van Hunen and van den Berg, 2008) and/or dehydration-induced stiffening and larger resistance to slab bending (Conrad and Hager, 2001; Korenaga, 2003, 2006). These stabilizing effects, however, have recently been questioned either because of greater upper-mantle depletion in the Archean that limited melting (Davies, 2006, 2007b) or because plate bending provides only minor resistance in convective circulation (Buffett and Becker, 2012; Davies, 2009; Leng and Zhong, 2010; see also **Section 7.07.5.5.3** and **Chapter 7.09**).

In the end, future plate generation theories that allow for thermal and petrologic evolution and inherited plate boundary weakness will be necessary to understand tectonic evolution and history.

7.07.7.2 Plates, Continents, and the Wilson Cycle

A major phase of petrologic evolution on Earth was the growth of continental crust through the Precambrian, which played a major role in the evolution of the plate–mantle system. In particular, with fully evolved continental crust, the dynamics of the mantle and lithosphere is influenced by the contrast between continents and oceans in myriad ways. First, while the ocean floor is continuously being created and destroyed, continents are essentially passive and permanent rafts floating at the surface of the Earth. In terms of basic physics, the continents impose a heterogeneous thermal boundary condition at the top of the mantle, which can significantly affect convective circulation (Kelly and Pal, 1978). Because of their thicker lithosphere, the heat flow across continents is around 65 mW m^{-2}, with values as low as 15 mW m^{-2} over Precambrian areas (Pinet et al., 1991). These heat flow values include the large radiogenic contribution in continental crust, which suggests that the heat flow across the Moho in cratonic lithosphere is roughly an order of magnitude smaller than

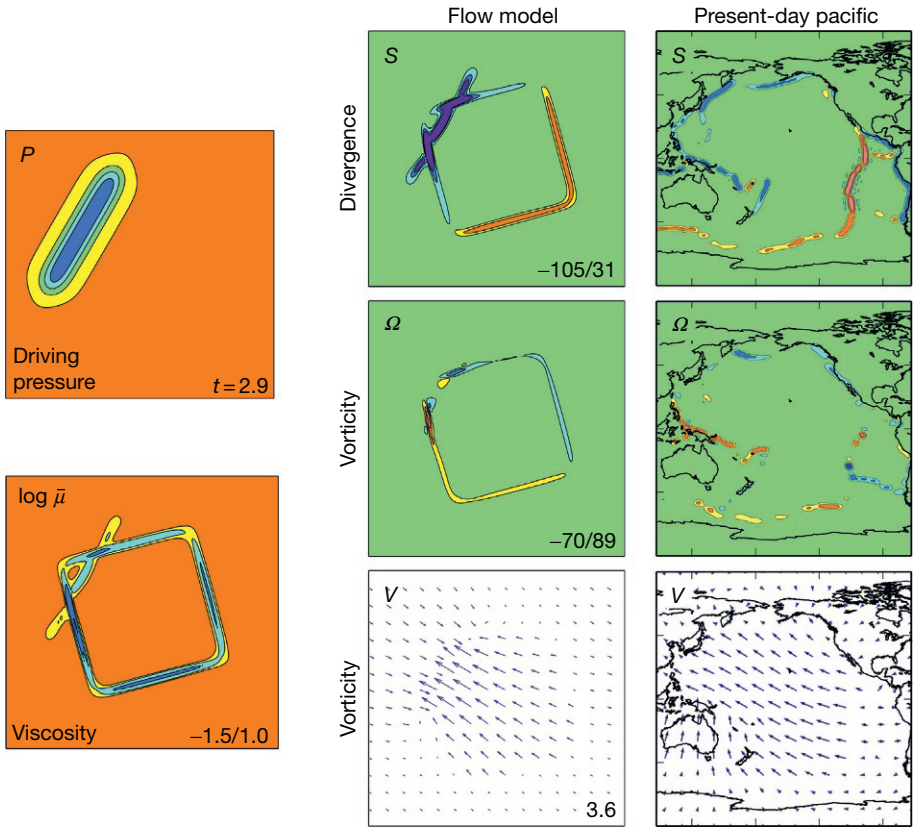


Figure 32 A similar model calculation to that shown in **Figure 31**, but now an abstraction of the rotation of the Pacific Plate associated with the Emperor–Hawaiian bend 47 Ma. The initial condition is a platelike flow similar to that shown in **Figure 31** but with velocity oriented along azimuth W15N, similar to the Emperor seamount track. The driving pressure is then rotated 45 degrees to drive flow W60N, similar to the Hawaiian Islands track, which is the final state shown. The weakened damaged zones that determined the configuration of the plate before the rotation are inherited and maintained after rotation and create oblique subduction, divergence, and strike-slip boundaries, in addition to spalling of minor plates. Although the calculation is highly idealized, it has significant resemblance to the structure of the current-day Pacific Plate and minor Philippine Plate (right column). (Extrema of contours or vectors are indicated on each frame.) Adapted from Bercovici D and Ricard Y (2014) Plate tectonics, damage and inheritance. *Nature* 506: 513–516. http://dx.doi.org/10.1038/nature13072.

that across the seafloor, which is on average around 100 mW m^{-2}. This heterogeneity in heat flow has various consequences for plate tectonics.

Because the thick continental lithosphere acts as a partial insulator, the underlying mantle tends to be hotter. This thermal blanketing has been observed experimentally and numerically (Guillou and Jaupart, 1995; Lenardic et al., 2011). The temperature increase depends on the size of the continents (Grigne et al., 2007; Phillips and Coltice, 2010) and could reach anomalies of 100 K under supercontinents (Coltice et al., 2007). This effect causes large-scale convective instabilities, which in turn induce rapid drift of continental rafts and subsequent supercontinent reorganization. This mantle–continent interaction provides a mechanism for the Wilson cycle – wherein continental plates aggregate over cold downwellings, inhibit subduction, overheat the underlying mantle, and then fragment over the resulting hot upwelling mantle – and has been modeled by Gurnis (1988) with simple 2-D Cartesian geometry, as well as with 3-D spherical geometry with platelike rheology (e.g., Rolf et al., 2012). According to the models and scaling analysis of Zhong and Gurnis (1993), the periodicity of this cycle, 300–500 My, corresponds to observed Wilson cycle periods. The instability and dispersal of supercontinents above a hotter mantle may also be reinforced when the subcontinental mantle is isolated by a sheet of subducting slabs (Heron and Lowman, 2011; Yoshida, 2013). However, as discussed in Section 7.07.5.3, recent work suggests the supercontinent insulation effect may be insufficient to drive continent dispersal, which thus remains an open question.

The large contrast in strength between oceanic and continental lithospheres also tends to localize stresses at their junction and facilitate development of weak zones and new plate boundaries. Convection models with plastic lithospheres (see **Section 7.07.6.1**) allow for such weakening at larger yield stresses with continents present than without them; this allows for better agreement with the experimental values of yield stress (Rolf and Tackley, 2011). Although the effects of continent/ocean heterogeneities have not been considered in damage models, a similar weakening and localization near heterogeneities can be expected. The presence of continents likely also affects the seafloor age–area distribution (see **Section 7.07.2.1.3**). In particular, without continents, subduction ideally happens once it reaches a critical age (i.e., it is old,

cold, and heavy enough to sink); but the imposition of continents could cause subduction to occur preferentially at passive margins that facilitated weakening and thus could be independent of seafloor age, which is in agreement with observations of present-day Earth (Becker et al., 2009; Coltice et al., 2012).

7.07.7.3 Plates, Climate, and Divergent Evolution of Planets

That Earth has plate tectonics but her ostensible twin Venus does not has been one of the key mysteries in the plate generation problem, and it has motivated much speculation about planetary conditions, including the requirement of liquid water, for plate tectonics to exist. While a traditional view has been that water lubricates plates by, for example, introduction of sediments at subduction zones or serpentinization along faults (e.g., Hilairet et al., 2007; Korenaga, 2007; Lenardic and Kaula, 1994; Tozer, 1985), the Earth's lithosphere might be as dry as that of Venus, because of dehydration melting at ridges (Hirth and Kohlstedt, 1996; see also **Section 7.07.6.2.2**). This paradox has prompted arguments that water's role is not in weakening the lithosphere but is instead in keeping Earth's surface temperature cool. In particular, Lenardic et al. (2008) posited that cooler surface temperatures on Earth lead to a more negatively buoyant top thermal boundary layer than on Venus; thus, Earth's lithospheric stress state is able to exceed its yield stress and fail, while Venus's lithosphere cannot. Alternatively, Landuyt and Bercovici (2009b) argued that surface temperature controls the competition between shear-localizing damage and healing (i.e., grain growth); that is, a cool surface retards healing and promotes damage and plate boundary formation, while a hot surface does the opposite and yields a stagnant lithosphere (**Figure 33**). These studies both suggest that Earth has plate tectonics because water drives the carbon cycle, which leads to a cool surface and the right conditions for plate tectonics, while Venus's loss of water and runaway greenhouse inhibits plate formation. However, this finding leads to the even deeper question of whether plate tectonics (which stabilizes a temperate climate) or a temperate climate (which is a necessary condition for tectonics) came first. Very recent coupled mantle-climate models suggest that the coupled feedbacks between tectonics, degassing, weathering, and greenhouse evolution occur simultaneously and can account for the divergent evolution between Earth and Venus (Driscoll and Bercovici, 2013).

7.07.7.4 Plates, Super-Earths, and Habitability

Plate tectonics is presumed to be a necessary condition for a temperate climate, by the carbon buffer and negative feedbacks associated with resurfacing and mineral exposure, erosion, weathering, and volcanism (Berner, 2004; Walker et al., 1981); the reciprocal dependence of plate tectonics on climate (**Section 7.07.7.3**) thus implies that a clement, habitable surface, and plate tectonics are mutually required. Plate tectonics may also be necessary for the existence of life by providing a source of thermodynamic disequilibrium through continuous recycling of the surface (e.g., Martin et al., 2008; Southam and Westall, 2007). The discovery of many terrestrial planets in other solar systems over the last 15 years (e.g., Charbonneau et al., 2009) has, therefore, emphasized the importance of understanding the conditions for plate tectonics as one

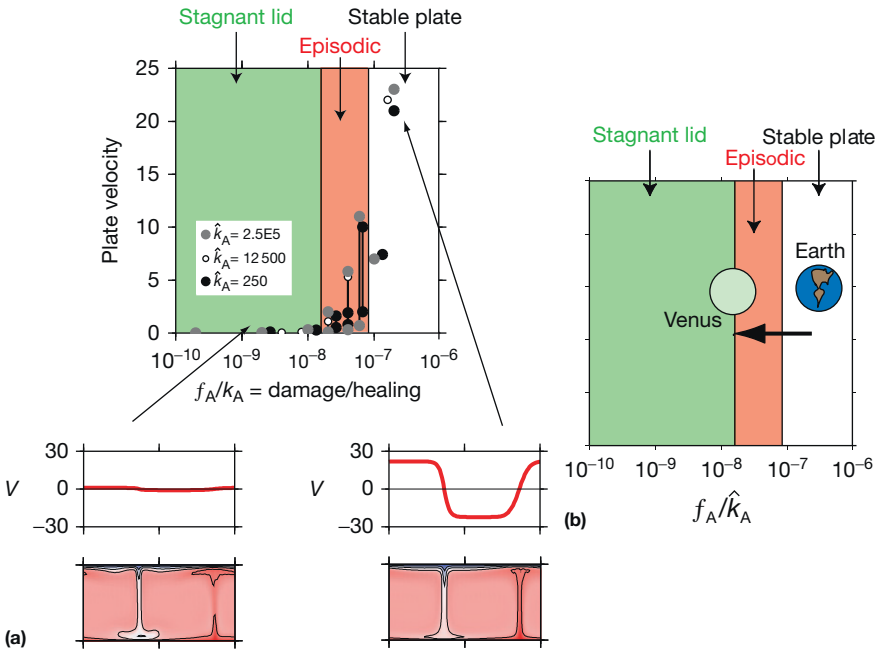


Figure 33 Convection with damage – transition from stagnant, to episodic, to platelike as a function of the damage–healing ratio (a). Healing is a function of surface temperature, and thus, high temperatures enhance healing relative to damage, while low surface temperature is opposite. This ostensibly explains why Venus is stagnant or episodic and Earth has plate tectonics (b). Adapted from Landuyt W and Bercovici D (2009) Variations in planetary convection via the effect of climate on damage. *Earth and Planetary Science Letters* 277: 29–37.

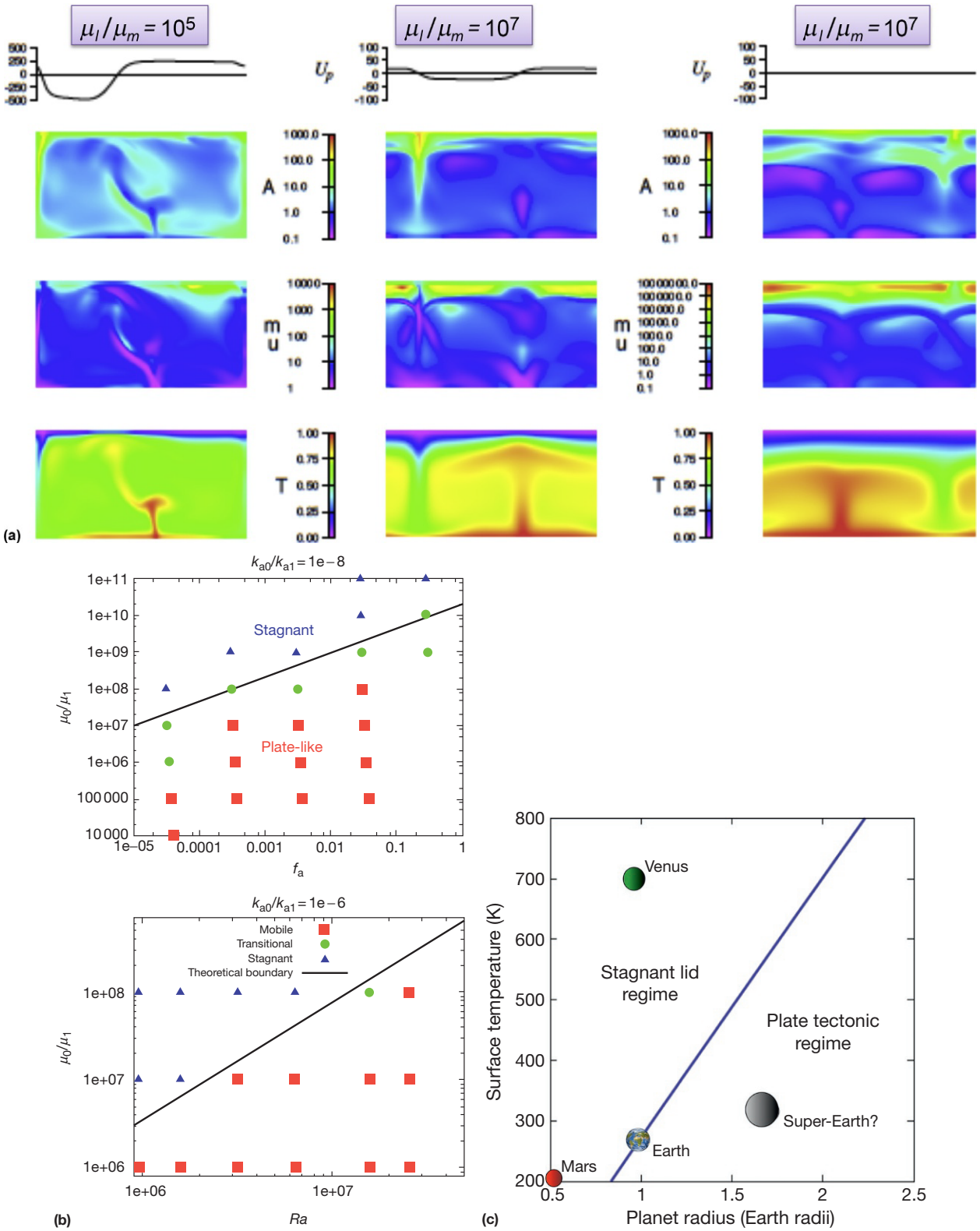


Figure 34 Convection with damage: the transition from stagnant to platelike or mobile surface flow occurs at a critical lithosphere–mantle viscosity ratio μ_l/μ_m (a). This ratio is itself a function of damage (for a given healing) and Rayleigh number (b) for which a scaling law for the transition can be inferred analytically. This can be compared to the predicted viscosity ratio of a planet with a certain size and internal and surface temperatures. Equivalence of these viscosity ratios occurs at a unique surface temperature for a given planet size (and internal temperature, not shown) and thus yields a phase diagram for planetary plate tectonics (c). Adapted from Foley BJ, Bercovici D, and Landuyt W (2012) The conditions for plate tectonics on super-earths: Inferences from convection models with damage. *Earth and Planetary Science Letters* 331–332: 281–290.

(although perhaps not a unique) requirement for liquid water and presumed habitability. The most readily available observation of terrestrial exoplanets is their mass, which has caused some debate as to whether size is more or less conducive to plate tectonics. Indeed, the first studies of this issue came to opposite conclusions, that is, plate tectonics on super-Earths is either inevitable (Valencia and O'Connell, 2009; Valencia et al., 2007) or impossible (O'Neill and Lenardic, 2007). More recent studies suggest that other factors such as surface temperature are equally or more important (Foley et al., 2012; Korenaga, 2010; van Heck and Tackley, 2011), which corresponds more closely to the Earth–Venus–Mars comparison (**Figure 34**). Overall, it appears that planet size is possibly more conducive to tectonics, depending on the surface conditions (Foley et al., 2012; Korenaga, 2010), interior composition and rheology (Tackley et al., 2013), and age in the planets thermal history (Lenardic and Crowley, 2012). However, this debate has highlighted the importance of understanding the physics of plate generation in order to predict how planetary conditions facilitate or inhibit plate-tectonic formation.

7.07.8 Closing and Future Directions

Although the theory of plate tectonics has provided one of the greatest predictive tools in the solid-Earth sciences, it is by no means a complete theory. Plate tectonics is a kinematic model in that it describes motions, but it does not involve dynamics, that is, either the forces or energy behind plate motions; nor does it explain the cause and generation of the plates themselves. In contrast, mantle convection theory describes and incorporates the essential dynamics and energy source for plate motions, but does not in itself predict the formation of the plates – with weak boundaries, strong plate interiors, asymmetrical subduction, passive spreading centers, sudden and sometimes global changes in plate motions, and toroidal motion. A theory of mantle convection that generates plates – in essence a unified theory of mantle dynamics and plate tectonics – remains one of the grand challenges of geodynamics.

Considerable progress has been made in the last 20 years on obtaining fluid mantle flow models that yield platelike behavior. The earlier successes come about by prescribing lithospheric weak zones and ad hoc rheologies (or rheologies like brittle failure that, while empirically based, are not applicable to the entire lithosphere) that yield more or less the correct behavior. While these ad hoc formulations gave important clues about what basic stress–strain rate laws allow platelike behavior, they did not elucidate the underlying physics leading to these rheologies or to plate generation.

More recent studies, in the last decade or so, have expanded our understanding of the physical and microscopic mechanisms behind shear localization, the depths and conditions at which these mechanisms occur, and their interaction with convective circulation. The areas of physics and chemistry involving melting, the role of water, fracture and void nucleation, and grain evolution are few examples of plate generation mechanisms that have been recently elucidated. However, much work remains to understand these processes, including how the thermal, chemical, and petrologic history of the Earth and other planets interacts with plate generation mechanisms to control tectonic and hence planet-system evolution. Moreover, observational, in particular seismic, data on the material properties and 3-D structure of plate boundaries (including ancient, dormant, active, and incipient ones) are of vital importance in constraining dynamic models and microscopic theories of plate boundary formation; for example, grain-size reduction or 'mylonitic' models of plate boundary formation are testable since they predict lack of lattice-preferred orientation and seismic anisotropy (which only occurs in conjunction with dislocation creep). As seismic coverage of the Earth continues to improve with the deployment of continental arrays (e.g., USArray) and ocean-bottom seismometers, along with more advanced technology such as receiver function techniques, the wealth of information will increase tremendously.

The theories of plate tectonics and mantle convection are far from being tried, true, and well-resolved models; indeed, they are still in their youth, and their successful unification will require years of ongoing research and advances in microscopic physics, continuum mechanics, and observational geophysics.

References

Abbott PL (2011) *Natural Disasters*, 8th edn. New York: McGraw-Hill.

Ahmed O and Lenardic A (2010) Low viscosity channels and the stability of long wavelength convection. *Physics of the Earth and Planetary Interiors* 179(3–4): 122–126.

Ahrens TJ and Schubert G (1975a) Gabbro-eclogite reaction rate and its geophysical significance. *Reviews of Geophysics and Space Physics* 13: 383–400.

Ahrens TJ and Schubert G (1975b) Rapid formation of eclogite in a slightly wet mantle. *Earth and Planetary Science Letters* 27: 90–94.

Allégre CJ (1982) Chemical geodynamics. *Tectonophysics* 81(3–4): 109–132.

Argus DF and Gordon RG (1991) No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1. *Geophysical Research Letters* 18(11): 2039–2042.

Balachandar S, Yuen D, and Reuteler D (1995a) Localization of toroidal motion and shear heating in 3-D high Rayleigh number convection with temperature-dependent viscosity. *Geophysical Research Letters* 22: 477–480.

Balachandar S, Yuen D, Reuteler D, and Lauer G (1995b) Viscous dissipation in 3-dimensional convection with temperature-dependent viscosity. *Science* 267: 1150–1152.

Beaumont C (1976) The evolution of sedimentary basins on a viscoelastic lithosphere. *Geophysical Journal of the Royal Astronomical Society* 55: 471–497.

Becker T, Conrad C, Buffett B, and Müller R (2009) Past and present seafloor age distributions and the temporal evolution of plate tectonic heat transport. *Earth and Planetary Science Letters* 278: 233–242.

Bénard H (1900) Les tourbillons cellulaires dans une nappe liquide. *Revue générale des Sciences pures et appliquées* 11: 1261–1271, and 1309–1328.

Bénard H (1901) Les tourbillons cellulaires dans une nappe liquide transportant de la chaleur par convection en régime permanent. *Annales de Chimie et de Physique* 23: 62–144.

Bercovici D (1993) A simple model of plate generation from mantle flow. *Geophysical Journal International* 114: 635–650.

Bercovici D (1995a) On the purpose of toroidal flow in a convecting mantle. *Geophysical Research Letters* 22: 3107–3110.

Bercovici D (1995b) A source–sink model of the generation of plate tectonics from non-Newtonian mantle flow. *Journal of Geophysical Research* 100: 2013–2030.

Bercovici D (1996) Plate generation in a simple model of lithosphere–mantle flow with dynamic self-lubrication. *Earth and Planetary Science Letters* 144: 41–51.

Bercovici D (1998) Generation of plate tectonics from lithosphere-mantle flow and void-volatile self-lubrication. *Earth and Planetary Science Letters* 154: 139–151.

Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121.

Bercovici D (2007) Mantle dynamics, past, present and future: An overview. In: Bercovici D and Schubert G (Chief Editor) (eds.) *Treatise on Geophysics*. *Mantle Dynamics*, vol. 7, pp. 1–30. New York: Elsevier.

Bercovici D and Karato S (2003) Theoretical analysis of shear localization in the lithosphere. In: Karato S and Wenk H (eds.) *Reviews in Mineralogy and Geochemistry: Plastic Deformation of Minerals and Rocks*, vol. 51, pp. 387–420. Washington, DC: Mineralogical Society of America (ch. 13).

Bercovici D and Kelly A (1997) Nonlinear initiation of diapirs and plume heads. *Physics of the Earth and Planetary Interiors* 101: 119–130.

Bercovici D and Long MD (2014) Slab rollback instability and supercontinent dispersal. *Geophysical Research Letters* 41. http://dx.doi.org/10.1002/2014GL061251.

Bercovici D and Ricard Y (2003) Energetics of a two-phase model of lithospheric damage, shear localization and plate-boundary formation. *Geophysical Journal International* 152: 581–596.

Bercovici D and Ricard Y (2005) Tectonic plate generation and two-phase damage: Void growth versus grainsize reduction. *Journal of Geophysical Research* 110: B03401. http://dx.doi.org/10.1029/2004JB003181.

Bercovici D and Ricard Y (2012) Mechanisms for the generation of plate tectonics by two-phase grain-damage and pinning. *Physics of the Earth and Planetary Interiors* 202–203: 27–55.

Bercovici D and Ricard Y (2013) Generation of plate tectonics with two-phase grain-damage and pinning: Source–sink model and toroidal flow. *Earth and Planetary Science Letters* 365: 275–288.

Bercovici D and Ricard Y (2014) Plate tectonics, damage and inheritance. *Nature* 508: 513–516. http://dx.doi.org/10.1038/nature13072.

Bercovici D, Ricard Y, and Richards M (2000) The relation between mantle dynamics and plate tectonics: A primer. In: Richards MA, Gordon R, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*. *Geophysical Monograph Series*, vol. 121, pp. 5–46. Washington, DC: American Geophysical Union.

Bercovici D, Ricard Y, and Schubert G (2001a) A two-phase model of compaction and damage: 1. General theory. *Journal of Geophysical Research* 106(B5): 8887–8906.

Bercovici D, Ricard Y, and Schubert G (2001b) A two-phase model of compaction and damage: 3. Applications to shear localization and plate boundary formation. *Journal of Geophysical Research* 106(B5): 8925–8940.

Bercovici D, Schubert G, and Glatzmaier G (1989a) Three-dimensional, spherical models of convection in the earth's mantle. *Science* 244: 950–955.

Bercovici D, Schubert G, and Glatzmaier G (1992) Three-dimensional, infinite Prandtl number, compressible convection in a basally heated spherical shell. *Journal of Fluid Mechanics* 239: 683–719.

Bercovici D, Schubert G, Glatzmaier G, and Zebib A (1989b) Three-dimensional thermal convection in a spherical shell. *Journal of Fluid Mechanics* 206: 75–104.

Bercovici D, Schubert G, and Ricard Y (2015) Abrupt tectonics and rapid slab detachment with grain damage. *Proceedings of the National Academy of Sciences*, http://dx.doi.org/10.1073/pnas.1415473112.

Berner R (2004) *The Phanerozoic Carbon Cycle*. Oxford: Oxford University Press.

Bird P (2003) An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems* 4(3): 1027. http://dx.doi.org/10.1029/2001GC000252.

Block M (1956) Surface tension as the cause for Bénard and surface deformation in a liquid film. *Nature* 178: 650–651.

Brady J (1995) Diffusion data for silicate minerals, glasses, and liquids. In: *Mineral Physics and Crystallography: A Handbook of Physical Constants*. *AGU Reference Shelf*, vol. 2, pp. 269–290. Washington, DC: American Geophysical Union.

Braun J, Chery J, Poliakov A, et al. (1999) A simple parameterization of strain localization in the ductile regime due to grain size reduction: A case study for olivine. *Journal of Geophysical Research* 104: 25167–25181.

Bridgman PW (1922) *Dimensional Analysis*. New Haven, CT: Yale University Press.

Brown M (2007) Metamorphic conditions in orogenic belts: A record of secular change. *International Geology Review* 49: 193–234.

Buffett BA and Becker TW (2012) Bending stress and dissipation in subducted lithosphere. *Journal of Geophysical Research – Solid Earth* 117. http.//dxdoi.org/10.1029/2012JB009205.

Bunge H, Richards M, and Baumgardner J (1996) Effect of depth-dependent viscosity on the planform of mantle-convection. *Nature* 379: 436–438.

Bunge H, Richards M, and Baumgardner J (1997) A sensitivity study of 3-D spherical mantle convection at 10^8 Rayleigh number: Effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *Journal of Geophysical Research* 102: 11991–12007.

Busse F (1975) Patterns of convection in spherical shells. *Journal of Fluid Mechanics* 72: 67–85.

Busse F (1978) Non-linear properties of thermal convection. *Reports on Progress in Physics* 41: 1929–1967.

Busse F and Riahi N (1982) Patterns of convection in spherical shells: Part 2. *Journal of Fluid Mechanics* 123: 283–301.

Busse F and Riahi N (1988) Mixed-mode patterns of bifurcations from spherically symmetric basic states. *Nonlinearity* 1: 379–388.

Busse F and Whitehead J (1971) Instabilities of convection rolls in high Prandtl number fluid. *Journal of Fluid Mechanics* 47: 305–320.

Cadek O and Ricard Y (1992) Toroidal/poloidal energy partitioning and global lithospheric rotation during cenozoic time. *Earth and Planetary Science Letters* 109: 621–632.

Cadek O, Ricard Y, Martinec Z, and Matyska C (1993) Comparison between Newtonian and non-Newtonian flow driven by internal loads. *Geophysical Journal International* 112: 103–114.

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. New York: Oxford University Press.

Charbonneau D, Berta ZK, Irwin J, et al. (2009) A super-earth transiting a nearby low-mass star. *Nature* 462: 891–894.

Choblet G, Cadek O, Couturier F, and Dumoulin C (2007) OEDIPUS: A new tool to study the dynamics of planetary interiors. *Geophysical Journal International* 170(1): 9–30. http://dx.doi.org/10.1111/j.1365-246X.2007.03419.x.

Christensen U (1984a) Convection with pressure- and temperature-dependent non-Newtonian rheology. *Geophysical Journal of the Royal Astronomical Society* 77: 343–384.

Christensen U (1984b) Heat transport by variable-viscosity convection and implications for the earth's thermal evolution. *Physics of the Earth and Planetary Interiors* 35: 264–282.

Christensen U (1984c) Instability of a hot thermal boundary layer and initiation of thermo-chemical plumes. *Annales Geophysicae* 2: 311–314.

Christensen U and Harder H (1991) Three-dimensional convection with variable viscosity. *Geophysical Journal International* 104: 213–226.

Christensen U and Yuen D (1985) Layered convection induced by phase changes. *Journal of Geophysical Research* 90: 291–300.

Coltice N, Bertrand H, Rey P, Jourdan F, Phillips BR, and Ricard Y (2009) Global warming of the mantle beneath continents back to the Archaean. *Gondwana Research* 15(3–4, SI): 254–266.

Coltice N, Phillips BR, Bertrand H, Ricard Y, and Rey P (2007) Global warming of the mantle at the origin of flood basalts over supercontinents. *Geology* 35(5): 391–394.

Coltice N, Rolf T, Tackley PJ, and Labrosse S (2012) Dynamic causes of the relation between area and age of the ocean floor. *Science* 336(6079): 335–338. http://www.sciencemag.org/content/336/6079/335.abstract.

Condie K and Kroner A (2008) When did plate tectonics begin? Evidence from the geologic record. In: Condie K and Pease V (eds.) *When Did Plate Tectonics Begin on Planet Earth?* pp. 281–294. Boulder, CO: Geological Society of America, Special Paper 440.

Connerney JEP, Acuña MH, Ness NF, et al. (2005) Tectonic implications of mars crustal magnetism. *Proceedings of the National Academy of Sciences of the United States of America* 102(42): 14970–14975. http://www.pnas.org/content/102/42/14970.abstract.

Connerney JEP, Acuña MH, Wasilewski PJ, et al. (1999) Magnetic lineations in the ancient crust of mars. *Science* 284(5415): 794–798. http://www.sciencemag.org/content/284/5415/794.abstract.

Conrad C and Hager B (2001) Mantle convection with strong subduction zones. *Geophysical Journal International* 144: 271–288.

Corrieu V, Ricard Y, and Froidevaux C (1994) Radial viscosity of the earth deduced from mantle tomography. *Physics of the Earth and Planetary Interiors* 84: 3–13.

Crameri F, Tackley PJ, Meilick I, Gerya TV, and Kaus BJP (2012) A free plate surface and weak oceanic crust produce single-sided subduction on earth. *Geophysical Research Letters* 39(3: L03306). http://dx.doi.org/10.1029/2011GL050046.

Creager K and Jordan T (1984) Slab penetration into the lower mantle. *Journal of Geophysical Research* 89: 3031–3049.

Davies G (1980) Thermal histories of convective earth models and constraints on radiogenic heat production in the earth. *Journal of Geophysical Research* 85: 2517–2530.

Davies G (1988a) Ocean bathymetry and mantle convection: 1. Large-scale flow and hotspots. *Journal of Geophysical Research* 93: 10467–10480.

Davies G (1988b) Ocean bathymetry and mantle convection: 2. Small scale flow. *Journal of Geophysical Research* 93: 10481–10488.

Davies G (1988c) The role of the lithosphere in mantle convection. *Journal of Geophysical Research* 93: 10451–10466.

Davies G (1992) On the emergence of plate tectonics. *Geology* 20: 963–966.

Davies GF (2006) Gravitational depletion of the early earth's upper mantle and the viability of early plate tectonics. *Earth and Planetary Science Letters* 243(3–4): 376–382, http://www.sciencedirect.com/science/article/pii/S0012821X06001051.

Davies G (2007a) Thermal evolution of the earth. In: Stevenson D and Schubert G (Chief Editor) (eds.) *Treatise on Geophysics*. *Evolution of the Earth*, vol. 9, pp. 197–216. Amsterdam: Elsevier.

Davies GF (2007b) Controls on density stratification in the early mantle. *Geochemistry, Geophysics, Geosystems* 8(4). http://dx.doi.org/10.1029/2006GC001414.

Davies GF (2009) Effect of plate bending on the Urey ratio and the thermal evolution of the mantle. *Earth and Planetary Science Letters* 287(3–4): 513–518. http://www.sciencedirect.com/science/article/pii/S0012821X09005238.

Davies G and Richards M (1992) Mantle convection. *Journal of Geology* 100: 151–206.

De Bresser J, Peach C, Reijs J, and Spiers C (1998) On dynamic recrystallization during solid state flow: Effects of stress and temperature. *Geophysical Research Letters* 25: 3457–3460.

De Bresser J, ter Heege J, and Spiers C (2001) Grain size reduction by dynamic recrystallization: Can it result in major rheological weakening? *International Journal of Earth Sciences* 90: 28–45.

Derby B and Ashby M (1987) On dynamic recrystallization. *Scripta Metallurgica* 21: 879–884.

Dijkstra AH, Drury MR, Vissers RLM, Newman J, and Van Roermund HLM (2004) Shear zones in the upper mantle: Evidence from alpine- and ophiolite-type peridotite massifs. *Geological Society, London, Special Publications* 224(1): 11–24. http://sp.lyellcollection.org/content/224/1/11.abstract.

Doherty RD, Hughes DA, Humphreys FJ, et al. (1997) Current issues in recrystallization: A review. *Materials Science and Engineering A* 238(2): 219–274.

Driscoll P and Bercovici D (2013) Divergent evolution of Earth and Venus: Influence of degassing, tectonics, and magnetic fields. *Icarus* 226: 1447–1464. http://www.sciencedirect.com/science/article/pii/S0019103513003266.

Duffield W (1972) A naturally occurring model of global plate tectonics. *Journal of Geophysical Research* 77: 2543–2555.

Dumoulin C, Bercovici D, and Wessel P (1998) A continuous plate-tectonic model using geophysical data to estimate plate margin widths, with a seismicity based example. *Geophysical Journal International* 133: 379–389.

Duncan R and Richards M (1991) Hotspots, mantle plumes, flood basalts and true polar wander. *Reviews of Geophysics* 29: 31–50.

Escartin J, Hirth G, and Evans B (2001) Strength of slightly serpentinized peridotites: Implications for the tectonics of oceanic lithosphere. *Geology* 29: 1023–1026.

Etheridge M and Wilkie J (1979) Grainsize reduction, grain boundary sliding and the flow strength of mylonites. *Tectonophysics* 58(1–2): 159–178.

Evans DA (2009) The palaeomagnetically viable, long-lived and all-inclusive Rodinia supercontinent reconstruction. *Geological Society, London, Special Publications* 327: 371–404.

Evans B and Kohlstedt D (1995) Rheology of rocks. In: Ahrens TJ (ed.) *Rock Physics and Phase Relations: A Handbook of Physical Constants*. *AGU Reference Shelf*, vol. 3, pp. 148–165. Washington, DC: American Geophysical Union.

Evans DA and Mitchell RN (2011) Assembly and breakup of the core of Paleoproterozoic–Mesoproterozoic supercontinent Nuna. *Geology* 39(5): 443–446. http://geology.gsapubs.org/content/39/5/443.abstract.

Evans DA and Pisarevsky S (2008) Plate tectonics on the early earth? Weighing the paleomagnetic evidence. In: Condie K and Pease V (eds.) *When Did Plate Tectonics Begin on Planet Earth?* pp. 249–263. Boulder, CO: Geological Society of America, Special Paper 440.

Faccenda M, Gerya TV, and Burlini L (2009) Deep slab hydration induced by bending-related variations in tectonic pressure. *Nature Geoscience* 2(11): 790–793. http://dx.doi.org/10.1038/ngeo656.

Faccenda M, Gerya TV, Mancktelow NS, and Moresi L (2012) Fluid flow during slab unbending and dehydration: Implications for intermediate-depth seismicity, slab weakening and deep water recycling. *Geochemistry, Geophysics, Geosystems* 13: Q01010.

Fleitout L and Froidevaux C (1980) Thermal and mechanical evolution of shear zones. *Journal of Structural Geology* 2: 159–164.

Foley B and Becker T (2009) Generation of plate-like behavior and mantle heterogeneity from a spherical, visco-plastic convection model. *Geochemistry, Geophysics, Geosystems* 10(8): Q08001. http://dx.doi.org/10.1029/2009GC002378.

Foley BJ, Bercovici D, and Landuyt W (2012) The conditions for plate tectonics on super-earths: Inferences from convection models with damage. *Earth and Planetary Science Letters* 331–332: 281–290.

Forsyth D, Scheirer D, Webb S, et al. (1998) Imaging the deep seismic structure beneath a mid-ocean ridge: The melt experiment. *Science* 280: 1215–1218.

Forsyth D and Uyeda S (1975) On the relative importance of the driving forces of plate motion. *Geophysical Journal of the Royal Astronomical Society* 43: 163–200.

Forte A and Peltier W (1987) Plate tectonics and aspherical earth structure: The importance of poloidal–toroidal coupling. *Journal of Geophysical Research* 92: 3645–3679.

Fukao Y, Obayashi M, Inoue H, and Nenbai M (1992) Subducting slabs stagnant in the mantle transition zone. *Journal of Geophysical Research – Solid Earth* 97(B4): 4809–4822.

Furbish D (1997) *Fluid Physics in Geology*. New York: Oxford University Press.

Furusho M and Kanagawa K (1999) Reaction induced strain localization in a lherzolite mylonite from the Hidaka metamorphic belt of central Hokkaido, Japan. *Tectonophysics* 313: 411–432.

Gable C, O'Connell R, and Travis B (1991) Convection in three dimensions with surface plates: Generation of toroidal flow. *Journal of Geophysical Research* 96: 8391–8405.

Gerya T (2010) Dynamical instability produces transform faults at mid-ocean ridges. *Science* 329(5995): 1047–1050.

Gerya TV (2013) Three-dimensional thermomechanical modeling of oceanic spreading initiation and evolution. *Physics of the Earth and Planetary Interiors* 214: 35–52.http://www.sciencedirect.com/science/article/pii/S0031920112001835.

Gerya TV, Connolly JA, and Yuen DA (2008) Why is terrestrial subduction one-sided? *Geology* 36(1): 43–46.

Giannandrea E and Christensen U (1993) Variable viscosity convection experiments with a stress-free upper boundary and implications for the heat transport in the earth's mantle. *Physics of the Earth and Planetary Interiors* 78: 139–152.

Glatzmaier G (1988) Numerical simulations of mantle convection: Time-dependent, three-dimensional, compressible, spherical shell. *Geophysical and Astrophysical Fluid Dynamics* 43: 223–264.

Glatzmaier G, Schubert G, and Bercovici D (1990) Chaotic subduction-like downflows in a spherical model of convection in the earth's mantle. *Nature* 347: 274–277.

Gordon R, DeMets C, and Royer J (1998) Evidence for long-term diffuse deformation of the lithosphere of the equatorial Indian ocean. *Nature* 395: 370–374.

Grand S (1994) Mantle shear structure beneath the Americas and surrounding oceans. *Journal of Geophysical Research* 99: 11591–11621.

Grand S, van der Hilst R, and Widiyantoro S (1997) Global seismic tomography: A snapshot of convection in the earth. *GSA Today* 7: 1–7.

Griffith A (1921) The phenomenon of rupture and flow in solids. *Philosophical Transactions of the Royal Society of London, Series A* 221: 163–198.

Grigne C, Labrosse S, and Tackley PJ (2007) Convection under a lid of finite conductivity: Heat flux scaling and application to continents. *Journal of Geophysical Research – Solid Earth* 112(B8).

Guillou L and Jaupart C (1995) On the effect of continents on mantle convection. *Journal of Geophysical Research – Solid Earth* 100(B12): 24217–24238.

Guillou L, Mareschal J, Jaupart C, Gariépy C, Bienfait G, and Lapointe R (1994) Heat flow, gravity, and structure of the Abitibi belt, superior province, Canada: Implications for mantle heat flow. *Earth and Planetary Science Letters* 122: 103–123.

Gurnis M (1988) Large-scale mantle convection and the aggregation and dispersal of supercontinents. *Nature* 332(6166): 695–699.

Gurnis M and Hager B (1988) Controls of the structure of subducted slabs. *Nature* 335: 317–321.

Gurnis M, Hall C, and Lavier L (2004) Evolving force balance during incipient subduction. *Geochemistry, Geophysics, Geosystems* 5(7).

Gurnis M, Zhong S, and Toth J (2000) On the competing roles of fault reactivation and brittle failure in generating plate tectonics from mantle convection. In: Richards MA, Gordon R, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*. *Geophysical Monograph Series*, vol. 121, pp. 73–94. Washington, DC: American Geophysical Union.

Hager B (1984) Subducted slabs and the geoid: Constraints on mantle rheology and flow. *Journal of Geophysical Research* 89: 6003–6015.

Hager B and O'Connell R (1978) Subduction zone dip angles and flow driven by plate motion. *Tectonophysics* 50: 111–133.

Hager B and O'Connell R (1979) Kinematic models of large-scale flow in the earth's mantle. *Journal of Geophysical Research* 84: 1031–1048.

Hager B and O'Connell R (1981) A simple global model of plate dynamics and mantle convection. *Journal of Geophysical Research* 86: 4843–4867.

Hales A (1936) Convection currents in the Earth. *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement* 3: 372–379.

Hall CE, Gurnis M, Sdrolias M, Lavier LL, and Mueller RD (2003) Catastrophic initiation of subduction following forced convergence across fracture zones. *Earth and Planetary Science Letters* 212: 15–30.

Hallam A (1987) Alfred Wegener and the hypothesis of continental drift. In: *Scientific Genius and Creativity: Readings from 'Scientific American' (Reprinted from* Scientific American, *February, 1975.)*, pp. 77–85. New York: W.H. Freeman & Co.

Hamano K, Abe Y, and Genda H (2013) Emergence of two types of terrestrial planet on solidification of magma ocean. *Nature* 497(7451): 607–610. http://dx.doi.org/10.1038/nature12163.

Hamilton WB (1998) Archean tectonics and magmatism. *International Geology Review* 40: 1–39.

Hansen LN, Zimmerman ME, and Kohlstedt DL (2012) The influence of microstructure on deformation of olivine in the grain-boundary sliding regime. *Journal of Geophysical Research* 117: B09201.

Harrison TM, Blichert-Toft J, Muller W, Albarede F, Holden P, and Mojzsis SJ (2005) Heterogeneous Hadean Hafnium: Evidence of continental crust at 4.4 to 4.5 Ga. *Science* 310(5756): 1947–1950.

Heron PJ and Lowman JP (2011) The effects of supercontinent size and thermal insulation on the formation of mantle plumes. *Tectonophysics* 510(1–2): 28–38.

Herwegh M, Linckens J, Ebert A, Berger A, and Brodhag S (2011) The role of second phases for controlling microstructural evolution in polymineralic rocks: A review. *Journal of Structural Geology* 33(12): 1728–1750.

Hess H (1962) History of ocean basins. In: Engeln A, James H, and Leonard B (eds.) *Petrologic Studies – A Volume in Honor of A.F. Buddington*, pp. 599–620. New York: Geological Society America.

Hey RN, Johnson PD, Martinez F, et al. (1995) Plate boundary reorganization at a large-offset, rapidly propagating rift. *Nature* 378(6553): 167–170. http://dx.doi.org/10.1038/378167a0.

Hey R and Wilson D (1982) Propagating rift explanation for the tectonic evolution of the northeast pacific—The pseudo-movie. *Earth and Planetary Science Letters* 58(2): 167–184. http://www.sciencedirect.com/science/article/pii/0012821X82901923.

Hilairet N, Reynard B, Wang Y, et al. (2007) High-pressure creep of serpentine, interseismic deformation, and initiation of subduction. *Science* 318(5858): 1910–1913.

Hilde T, Uyeda S, and Kroenke L (1977) Evolution of the western pacific and its margins. *Tectonophysics* 38: 145–165.

Hirth G and Kohlstedt D (1996) Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144: 93–108.

Hirth G and Kohlstedt D (2003) Rheology of the upper mantle and the mantle wedge: A view from the experimentalists. In: Eiler J (ed.) *Subduction Factor Monograph*, vol. 138, pp. 83–105. Washington, DC: American Geophysical Union.

Hofmann A (1997) Mantle geochemistry: The message from oceanic volcanism. *Nature* 385: 219–228.

Holmes A (1931) Radioactivity and earth movements. *Geological Society of Glasgow, Transactions* 18: 559–606.

Honda S, Yuen D, Balachandar S, and Reuteler D (1993) Three-dimensional instabilities of mantle convection with multiple phase transitions. *Science* 259: 1308–1311.

Hopkins MD, Harrison TM, and Manning CE (2010) Constraints on hadean geodynamics from mineral inclusions in >4 Ga zircons. *Earth and Planetary Science Letters* 298(3–4): 367–376.

Huang J and Zhong S (2005) Sublithospheric small-scale convection and its implications for the residual topography at old ocean basins and the plate model. *Journal of Geophysical Research* 110: B05404.

Ildefonse B, Blackman D, John B, et al. (2007) Oceanic core complexes and crustal accretion at slow-spreading ridges. *Geology* 35(7): 623–626. http://geology.gsapubs.org/content/35/7/623.abstract.

Ito E, Akaogi M, Topor L, and Navrotsky A (1990) Negative pressure–temperature slopes for reactions forming $MgSiO_3$ perovskite from calorimetry. *Science* 249: 1275–1278.

Ito K and Kennedy GC (1971) An experimental study of the basalt-garnet granulite–eclogite transition. *The Structure and Physical Properties of the Earth's Crust*. *Geophysical Monograph*, vol. 14, pp. 303–314. Washington, DC: American Geophysical Union.

Jarvis G and McKenzie D (1980) Convection in a compressible fluid with infinite Prandtl number. *Journal of Fluid Mechanics* 96: 515–583.

Jaupart C, Labrosse S, and Mareschal J-C (2007) Temperatures, heat and energy in the mantle of the earth. In: Bercovici D and Schubert G (Chief Editor) (eds.) *Treatise on Geophysics*. *Mantle Dynamics*, vol. 7, pp. 253–303. New York: Elsevier.

Jellinek A and Manga M (2004) Links between long-lived hotspots, mantle plumes, D″ and plate tectonics. *Reviews of Geophysics* 42(3). http://dx.doi.org/10.1029/2003RG000144.

Jenner F, Bennett V, Yaxley G, Friend C, and Nebel O (2013) Eoarchean within-plate basalts from southwest Greenland. *Geology* 41(3): 327–330.

Jin D, Karato S, and Obata M (1998) Mechanisms of shear localization in the continental lithosphere: Inference from the deformation microstructures of peridotites from the Ivrea zone, northwestern Italy. *Journal of Structural Geology* 20: 195–209.

Kameyama M, Yuen D, and Fujimoto H (1997) The interaction of viscous heating with grain-size dependent rheology in the formation of localized slip zones. *Geophysical Research Letters* 24: 2523–2526.

Kaminski E and Ribe N (2001) A kinematic model for recrystallization and texture development in olivine polycrystals. *Earth and Planetary Science Letters* 189(3–4): 253–267. http://www.sciencedirect.com/science/article/pii/S0012821X01003569.

Karato S (1989) Grain growth kinetics in olivine aggregates. *Tectonophysics* 168: 255–273.

Karato S (2008) *Deformation of Earth Materials: An Introduction to the Rheology of Solid Earth*. New York: Cambridge University Press.

Karato S, Toriumi M, and Fujii T (1980) Dynamic recrystallization of olivine single crystals during high temperature creep. *Geophysical Research Letters* 7: 649–652.

Karato S and Wu P (1993) Rheology of the upper mantle: A synthesis. *Science* 260: 771–778.

Katz RF (2008) Magma dynamics with the enthalpy method: Benchmark solutions and magmatic focusing at mid-ocean ridges. *Journal of Petrology* 49(12): 2099–2121.

Kaula W (1980) Material properties for mantle convection consistent with observed surface fields. *Journal of Geophysical Research* 85: 7031–7044.

Kelly R and Pal D (1978) Thermal-convection with spatially periodic boundary-conditions – Resonant wavelength excitation. *Journal of Fluid Mechanics* 86: 433–456.

Kemp D and Stevenson D (1996) A tensile flexural model for the initiation of subduction. *Geophysical Journal International* 125: 73–94.

Kennet B and Engdahl E (1991) Travel times for global earthquake location and phase identification. *Geophysical Journal International* 105: 429–465.

King S (1995) The viscosity structure of the mantle. In: *Reviews of Geophysics (Supplement): U.S. Quadrennial Report to the IUGG 1991-1994*, pp. 11–17. Washington, DC: American Geophysical Union.

King S (2007) Mantle downwellings and the fate of subducting slabs: Constraints from seismology, geoid, topography, geochemistry, and petrology. In: Bercovici D and Schubert G (Chief Editor) (eds.) *Treatise on Geophysics*. *Mantle Dynamics*, vol. 7, pp. 325–370. New York: Elsevier.

King S, Gable C, and Weinstein S (1991) Models of convection driven plates: A comparison of methods and results. *Geophysical Journal International* 109: 481–487.

King S and Hager B (1990) The relationship between plate velocity and trench viscosity in Newtonian and power-law subduction calculations. *Geophysical Research Letters* 17: 2409–2412.

King SD, Lowman JP, and Gable CW (2002) Episodic tectonic plate reorganizations driven by mantle convection. *Earth and Planetary Science Letters* 203(1): 83–91.

Kohlstedt D, Evans B, and Mackwell S (1995) Strength of the lithosphere: Constraints imposed by laboratory experiments. *Journal of Geophysical Research* 100: 17587–17602.

Korenaga J (2003) Energetics of mantle convection and the fate of fossil heat. *Geophysical Research Letters* 30(8). http://dx.doi.org/10.1029/2003GL016982.

Korenaga J (2006) Archean geodynamics and the thermal evolution of earth. In: Benn K, Mareschal J-C, and Condie K (eds.) *Archean Geodynamics and Environments*. *AGU Geophysical Monograph Series*, vol. 164, pp. 7–32. Washington, DC: American Geophysical Union.

Korenaga J (2007) Thermal cracking and the deep hydration of oceanic lithosphere: A key to the generation of plate tectonics? *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2006JB004502.

Korenaga J (2008) Urey ratio and the structure and evolution of earth's mantle. *Reviews of Geophysics* 46: RG2007. http://dx.doi.org/10.1029/2007RG000241.

Korenaga J (2010) On the likelihood of plate tectonics on super-earths: Does size matter? *Astrophysical Journal Letters* 725(1): L43–L46.

Korenaga J (2013) Initiation and evolution of plate tectonics on earth: Theories and observations. *Annual Review of Earth and Planetary Sciences* 41: 117–151.

Labrosse S (2002) Hotspots, mantle plumes and core heat loss. *Earth and Planetary Science Letters* 199(1–2): 147–156. http://www.sciencedirect.com/science/article/pii/S0012821X0200537X.

Lachenbruch A (1976) Dynamics of a passive spreading center. *Journal of Geophysical Research* 81: 1883–1902.

Lachenbruch A and Sass J (1980) Heat flow and energetics of the San Andreas fault zone. *Technical Report 80-625*, U.S. Geological Survey.

Landuyt W and Bercovici D (2009a) Formation and structure of lithospheric shear zones with damage. *Physics of the Earth and Planetary Interiors* 175: 115–126.

Landuyt W and Bercovici D (2009b) Variations in planetary convection via the effect of climate on damage. *Earth and Planetary Science Letters* 277: 29–37.

Landuyt W, Bercovici D, and Ricard Y (2008) Plate generation and two-phase damage theory in a model of mantle convection. *Geophysical Journal International* 174: 1065–1080.

Lay T, Hernlund J, and Buffett BA (2008) Core–mantle boundary heat flow. *Nature Geoscience* 1(1): 25–32. http://dx.doi.org/10.1038/ngeo.2007.44.

Le Pichon X (1968) Sea-floor spreading and continental drift. *Journal of Geophysical Research* 73(12): 3661–3697. http://dx.doi.org/10.1029/JB073i012p03661.

Lebrun J-F, Lamarche G, and Collot J-Y (2003) Subduction initiation at a strike-slip plate boundary: The cenozoic pacific-Australian plate boundary, south of New Zealand. *Journal of Geophysical Research* 108: 2453. http://dx.doi.org/10.1029/2002JB002041, B9.

Lebrun T, Massol H, Chassefière E, et al. (2013) Thermal evolution of an early magma ocean in interaction with the atmosphere. *Journal of Geophysical Research, Planets* 118(6): 1155–1176. http://dx.doi.org/10.1002/jgre.20068.

Lee K, Jiang Z, and Karato S (2002) A scanning electron microscope study of effects of dynamic recrystallization on the lattice preferred orientation in olivine. *Tectonophysics* 351: 331–341.

Lee C-TA, Lenardic A, Cooper CM, Niu F, and Levander A (2005) The role of chemical boundary layers in regulating the thickness of continental and oceanic thermal boundary layers. *Earth and Planetary Science Letters* 230(3–4): 379–395, http://www.sciencedirect.com/science/article/pii/S0012821X04007022.

Lenardic A and Crowley JW (2012) On the notion of well-defined tectonic regimes for terrestrial planets in this solar system and others. *The Astrophysical Journal* 755(2): 132. http://stacks.iop.org/0004-637X/755/i=2/a=132.

Lenardic A, Jellinek M, and Moresi L-N (2008) A climate change induced transition in the tectonic style of a terrestrial planet. *Earth and Planetary Science Letters* 271: 34–42.

Lenardic A and Kaula W (1994) Self-lubricated mantle convection: Two-dimensional models. *Geophysical Research Letters* 21: 1707–1710.

Lenardic A and Kaula W (1996) Near surface thermal/chemical boundary layer convection at infinite Prandtl number: Two-dimensional numerical experiments. *Geophysical Journal International* 126: 689–711.

Lenardic A, Moresi LN, Jellinek AM, and Manga M (2005) Continental insulation, mantle cooling, and the surface area of oceans and continents. *Earth and Planetary Science Letters* 234(3–4): 317–333.

Lenardic A, Moresi L, Jellinek AM, O'Neill CJ, Cooper CM, and Lee CT (2011) Continents, supercontinents, mantle thermal mixing, and mantle thermal isolation: Theory, numerical simulations, and laboratory experiments. *Geochemistry, Geophysics, Geosystems* 12. http://dxdoi.org/10.1029/2011GC003663.

Lenardic A, Richards MA, and Busse FH (2006) Depth-dependent rheology and the horizontal length scale of mantle convection. *Journal of Geophysical Research* 111: B07404. http://dx.doi.org/10.1029/2005JB003639.

Leng W and Zhong S (2010) Constraints on viscous dissipation of plate bending from compressible mantle convection. *Earth and Planetary Science Letters* 297(1–2): 154–164. http://www.sciencedirect.com/science/article/pii/S0012821X10003900.

Lewis C (2002) Arthur Holmes: An ingenious geoscientist. *GSA Today* 12(3): 16–17.

Linckens J, Herwegh M, Müntener O, and Mercolli I (2011) Evolution of a polymineralic mantle shear zone and the role of second phases in the localization of deformation. *Journal of Geophysical Research* 116: B06210, 21 pp.

Lister CRB (1972) On the thermal balance of a mid-ocean ridge. *Geophysical Journal of the Royal Astronomical Society* 26(5): 515–535. http://dx.doi.org/10.1111/j.1365-246X.1972.tb05766.x.

Lithgow-Bertelloni C and Richards M (1995) Cenozoic plate driving forces. *Geophysical Research Letters* 22: 1317–1320.

Lithgow-Bertelloni C and Richards MA (1998) Dynamics of cenozoic and mesozoic plate motions. *Reviews of Geophysics* 36: 27–78.

Lithgow-Bertelloni C, Richards M, Ricard Y, O'Connell R, and Engebretson D (1993) Toroidal–poloidal partitioning of plate motions since 120 MA. *Geophysical Research Letters* 20: 375–378.

Lithgow-Bertelloni C and Silver P (1998) Dynamic topography, plate driving forces and the African superswell. *Nature* 395: 269–272.

Liu L (1979) Phase transformations and the constitution of the deep mantle. In: McElhinny M (ed.) *Structure and Evolution Earth: Its Origins*, pp. 177–202. San Diego, CA: Academic Press.

Long MD and Becker T (2010) Mantle dynamics and seismic anisotropy. *Earth and Planetary Science Letters* 297(3–4): 341–354.

Long MD and Silver PG (2008) The subduction zone flow field from seismic anisotropy: A global view. *Science* 319(5861): 315–318. http://www.sciencemag.org/content/319/5861/315.abstract.

Lowman J (2011) Mantle convection models featuring plate tectonic behavior: An overview of methods and progress. *Tectonophysics* 510(1): 1–16.

Lowman J and Jarvis G (1996) Continental collisions in wide aspect ratio and high Rayleigh number two-dimensional mantle convection models. *Journal of Geophysical Research* 101(B11): 25485–25497.

Lowman J, King S, and Gable C (2003) The role of the heating mode of the mantle in periodic reorganizations of the plate velocity field. *Geophysical Journal International* 152: 455–467.

Machetel P and Weber P (1991) Intermittent layered convection in a model mantle with an endothermic phase change at 670 km. *Nature* 350: 55–57.

Malvern L (1969) *Introduction to the Mechanics of a Continuous Medium*. Englewood Cliffs, NJ: Prentice-Hall.

Martin W, Baross J, Kelley D, and Russell MJ (2008) Hydrothermal vents and the origin of life. *Nature Reviews Microbiology* 6(11): 805–814.

McKenzie D (1981) The variation of temperature with time and hydrocarbon maturation in sedimentary basins formed by extension. *Earth and Planetary Science Letters* 55(1): 87–98.

McKenzie D and Parker R (1967) The north pacific: An example of tectonics on a sphere. *Nature* 216: 1276–1280.

Mehl L and Hirth G (2008) Plagioclase preferred orientation in layered mylonites: Evaluation of flow laws for the lower crust. *Journal of Geophysical Research* 113: B05202. http://dx.doi.org/10.1029/2007JB005075.

Mojzsis SJ, Harrison TM, and Pidgeon RT (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409(6817): 178–181. http://dx.doi.org/10.1038/35051557.

Molnar P and Stock J (1987) Relative motions of hotspots in the pacific, Atlantic and Indian oceans since late cretaceous time. *Nature* 327(6123): 587–591.

Monnereau M and Quéré S (2001) Spherical shell models of mantle convection with tectonic plates. *Earth and Planetary Science Letters* 184(3–4): 575–587. http://www.sciencedirect.com/science/article/pii/S0012821X00003344.

Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl ER, and Hung S-H (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303: 338–343.

Montési L and Hirth G (2003) Grain size evolution and the rheology of ductile shear zones: From laboratory experiments to postseismic creep. *Earth and Planetary Science Letters* 211: 97–110.

Moresi L and Solomatov V (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic style of the Earth and Venus. *Geophysical Journal International* 133: 669–682.

Morgan W (1968) Rises, trenches, great faults and crustal blocks. *Journal of Geophysical Research* 73: 1959–1982.

Morgan WJ (1972) Plate motions and deep mantle convection. *Memoir – Geological Society of America* 132: 7–22.

Morgan JP and Smith W (1992) Flattening of the seafloor depth-age curve as a response to asthenospheric flow? *Nature* 359: 524–527.

Morra G, Seton M, Quevedo L, and Müller RD (2013) Organization of the tectonic plates in the last 200 Myr. *Earth and Planetary Science Letters* 373: 93–101.

Moyen J-F and van Hunen J (2012) Short-term episodicity of Archaean plate, tectonics. *Geology*, http://dx.doi.org/10.1130/G32894.1.

Mueller S and Phillips R (1991) On the initiation of subduction. *Journal of Geophysical Research* 96: 651–665.

Nakakuki T, Hamada C, and Tagawa M (2008) Generation and driving forces of plate-like motion and asymmetric subduction in dynamical models of an integrated mantle–lithosphere system. *Physics of the Earth and Planetary Interiors* 166(3–4): 128–146. http://www.sciencedirect.com/science/article/pii/S0031920108000241.

Nataf H and VanDecar J (1993) Seismological detection of a mantle plume? *Nature* 364: 115–120.

Nicolis G (1995) *Introduction to Nonlinear Science*. Cambridge: Cambridge University Press.

Nikolaeva K, Gerya TV, and Marques FO (2010) Subduction initiation at passive margins: Numerical modeling. *Journal of Geophysical Research* 115: B03406.

O'Connell R, Gable C, and Hager B (1991) Toroidal–poloidal partitioning of lithospheric plate motion. In: Sabadini R, et al. (eds.) *Glacial Isostasy, Sea Level and Mantle Rheology*, pp. 535–551. Norwell, MA: Kluwer Academic.

Ogawa M, Schubert G, and Zebib A (1991) Numerical simulations of three-dimensional thermal convection in a fluid with strongly temperature-dependent viscosity. *Journal of Fluid Mechanics* 233: 299–328.

Olgaard D (1990) The role of second phase in localizing deformation. In: Knipe R and Rutter E (eds.) *Deformation Mechanisms, Rheology and Tectonics*, vol. 54, pp. 175–181. London: Geological Society, Special Publication.

Olson P and Bercovici D (1991) On the equipartitioning of kinetic energy in plate tectonics. *Geophysical Research Letters* 18: 1751–1754.

Olson P and Corcos G (1980) A boundary layer model for mantle convection with surface plates. *Geophysical Journal of the Royal Astronomical Society* 62: 195–219.

O'Neill C and Lenardic A (2007) Geological consequences of super-sized earths. *Geophysical Research Letters* 34: 19204–19208.
O'Rourke J and Korenaga J (2012) Terrestrial planet evolution in the stagnant-lid regime: Size effects and the formation of self-destabilizing crust. *Icarus* 221: 1043–1060.
Oxburgh E and Turcotte D (1978) Mechanisms of continental drift. *Reports on Progress in Physics* 41: 1249–1312.
Parmentier E, Turcotte D, and Torrance K (1976) Studies of finite amplitude non-Newtonian thermal convection with applications to convection in the earth's mantle. *Journal of Geophysical Research* 81: 1839–1846.
Parsons B and McKenzie D (1978) Mantle convection and the thermal structure of plates. *Journal of Geophysical Research* 83: 4485–4496.
Parsons B and Sclater J (1977) An analysis of the variation of ocean floor bathymetry with age. *Journal of Geophysical Research* 82: 803–827.
Pearson J (1958) On convection cells induced by surface tension. *Journal of Fluid Mechanics* 4: 489–500.
Pekeris C (1935) Thermal convection in the interior of the earth. *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement* 3: 343–367.
Phillips BR and Coltice N (2010) Temperature beneath continents as a function of continental cover and convective wavelength. *Journal of Geophysical Research – Solid Earth* 115: B04408. http://dx.doi.org/10.1029/2009JB006600.
Pinet C, Jaupart C, Mareschal J, Gariepy C, Bienfait G, and Lapointe R (1991) Heat-flow and structure of the lithosphere in the Eastern Canadian Shield. *Journal of Geophysical Research – Solid Earth* 96(B12): 19941–19963.
Polat A, Appel PW, and Fryer BJ (2011) An overview of the geochemistry of Eoarchean to Mesoarchean ultramafic to mafic volcanic rocks, SW Greenland: Implications for mantle depletion and petrogenetic processes at subduction zones in the early Earth. *Gondwana Research* 20(2–3): 255–283.
Pollack H, Hurter S, and Johnson J (1993) Heat flow from the earth's interior: Analysis of the global data set. *Reviews of Geophysics* 31: 267–280.
Ranalli G (1995) *Rheology of the Earth*. London: Chapman and Hall.
Ranero CR, Phipps Morgan J, McIntosh K, and Reichert C (2003) Bending-related faulting and mantle serpentinization at the middle America trench. *Nature* 425(6956): 367–373. http://dx.doi.org/10.1038/nature01961.
Ratcliff J, Tackley P, Schubert G, and Zebib A (1997) Transitions in thermal convection with strongly variable viscosity. *Physics of the Earth and Planetary Interiors* 102: 201–212.
Regenauer-Lieb K, Yuen D, and Branlund J (2001) The initiation of subduction: Criticality by addition of water? *Science* 294: 578–580.
Ribe N (1992) The dynamics of thin shells with variable viscosity and the origin of toroidal flow in the mantle. *Geophysical Journal International* 110: 537–552.
Ricard Y and Bercovici D (2003) Two-phase damage theory and crustal rock failure: The theoretical 'void' limit, and the prediction of experimental data. *Geophysical Journal International* 155: 1057–1064.
Ricard Y, Bercovici D, and Schubert G (2001) A two-phase model of compaction and damage: 2. Applications to compaction, deformation, and the role of interfacial surface tension. *Journal of Geophysical Research* 106(B5): 8907–8924.
Ricard Y, Doglioni C, and Sabadini R (1991) Differential rotation between lithosphere and mantle: A consequence of lateral viscosity variations. *Journal of Geophysical Research* 96: 8407–8415.
Ricard Y, Fleitout L, and Froidevaux C (1984) Geoid heights and lithospheric stresses for a dynamic earth. *Annales Geophysicae* 2: 267–286.
Ricard Y and Froidevaux C (1986) Stretching instabilities and lithospheric boudinage. *Journal of Geophysical Research* 91: 8314–8324.
Ricard Y and Froidevaux C (1991) Seismic imaging, plate velocities and geoid, the direct and inverse problem. In: Sabadini R and Lambeck K (eds.) *Glacial Isostasy, Sea-Level and Mantle Rheology*, pp. 553–569. Dordrecht: Kluwer Academic.
Ricard Y, Richards MA, Lithgow-Berteloni C, and Stunf YL (1993) A geodynamic model of mantle mass heterogeneities. *Journal of Geophysical Research* 98: 21895–21909.
Ricard Y and Vigny C (1989) Mantle dynamics with induced plate tectonics. *Journal of Geophysical Research* 94: 17543–17559.
Ricard Y, Vigny C, and Froidevaux C (1989) Mantle heterogeneities, geoid and plate motion: A Monte Carlo inversion. *Journal of Geophysical Research* 94: 13739–13754.
Ricard Y and Wuming B (1991) Inferring the mantle viscosity and its three dimensional structure from geoid, topography and plate velocities. *Geophysical Journal* 105: 561–571.
Richards MA and Engebretson DC (1992) Large-scale mantle convection and the history of subduction. *Nature* 355(6359): 437–440.
Richards M and Hager B (1984) Geoid anomalies in a dynamic earth. *Journal of Geophysical Research* 89: 5487–6002.
Richards M and Lithgow-Bertelloni C (1996) Plate motion changes, the Hawaiian–Emperor bend, and the apparent success and failure of geodynamical models. *Earth and Planetary Science Letters* 137: 19–27.
Richards M, Yang W-S, Baumgardner J, and Bunge H-P (2001) Role of a low-viscosity zone in stabilizing plate tectonics: Implications for comparative terrestrial planetology. *Geochemistry, Geophysics, Geosystems* 2: 1026. http://dx.doi.org/10.1029/2000GC000115.
Ringwood A (1972) Phase transformations and mantle dynamics. *Earth and Planetary Science Letters* 14(2): 233–241. http://www.sciencedirect.com/science/article/pii/0012821X7290012X.
Ringwood A and Green D (1966) An experimental investigation of the gabbro-eclogite transformation and some geophysical implications. *Tectonophysics* 3(5): 383–427.http://www.sciencedirect.com/science/article/pii/0040195166900096.
Rolf T, Coltice N, and Tackley PJ (2012) Linking continental drift, plate tectonics and the thermal state of the Earth's mantle. *Earth and Planetary Science Letters* 351: 134–146.
Rolf T and Tackley PJ (2011) Focussing of stress by continents in 3D spherical mantle convection with self-consistent plate tectonics. *Geophysical Research Letters* 38(18): L18301. http://dx.doi.org/10.1029/2011GL048677.
Rozel A, Ricard Y, and Bercovici D (2011) A thermodynamically self-consistent damage equation for grain size evolution during dynamic recrystallization. *Geophysical Journal International* 184(2): 719–728.
Runcorn S (1962a) Convection currents in the earth's mantle. *Nature* 195: 1248–1249.
Runcorn S (1962b) Towards a theory of continental drift. *Nature* 193: 311–314.
Russo R and Silver PG (1994) Trench-parallel flow beneath the Nazca plate from seismic anisotropy. *Science* 263: 1105–1111.
Schmeling S and Jacoby W (1981) On modeling the lithosphere in mantle convection with non-linear rheology. *Journal of Geophysics* 50: 89–100.
Schubert G (1980) Self-driven motions of plates and descending slabs. In: Davies P and Runcorn S (eds.) *Mechanisms of Continental Drift and Plate Tectonics*, pp. 151–158. New York: Academic Press.
Schubert G, Froidevaux C, and Yuen D (1976) Oceanic lithosphere and asthenosphere: Thermal and mechanical structure. *Journal of Geophysical Research* 81: 3525–3540.
Schubert G and Sandwell D (1995) A global survey of possible subduction sites on Venus. *Icarus* 117(1): 173–196.
Schubert G, Stevenson D, and Cassen P (1980) Whole planet cooling and the radiogenic heat source contents of the earth and moon. *Journal of Geophysical Research* 85: 2531–2538.
Schubert G and Turcotte D (1972) One-dimensional model of shallow mantle convection. *Journal of Geophysical Research* 77: 945–951.
Schubert G, Turcotte D, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.
Schubert G, Turcotte D, and Yuen D (1975) Role of phase transitions in a dynamic mantle. *Geophysical Journal of the Royal Astronomical Society* 42: 705–735.
Schubert G and Zhang K (1997) Foundering of the lithosphere at the onset of subduction. *Geophysical Research Letters* 24: 1527–1529.
Sclater J, Jaupart C, and Galson D (1980) The heat flow through the oceanic and continental crust and the heat loss of the earth. *Reviews of Geophysics and Space Physics* 18: 269–311.
Sharp WD and Clague DA (2006) 50-Ma Initiation of Hawaiian–Emperor bend records major change in Pacific plate motion. *Science* 313(5791): 1281–1284.
Shen Y, Solomon S, Bjarnason I, and Wolfe C (1998) Seismic evidence for a lower mantle origin of the iceland plume. *Nature* 395: 62–65.
Shimizu I (1998) Stress and temperature dependence of recrystallized grain size: A subgrain misorientation model. *Geophysical Research Letters* 25: 4237–4240.
Shirey S, Kamber B, Whitehouse M, Mueller P, and Basu A (2008) A review of the isotopic and trace element evidence for mantle and crustal processes in the Hadean and Archean: Implications for the onset of plate tectonic subduction. In: Condie K and Pease V (eds.) *When Did Plate Tectonics Begin on Planet Earth?* pp. 1–29. Boulder, CO: Geological Society of America, Special Paper 440.
Shirey SB and Richardson SH (2011) Start of the Wilson cycle at 3 Ga shown by diamonds from subcontinental mantle. *Science* 333: 434–436.
Silver P, Carlson R, and Olson P (1988) Deep slabs, geochemical heterogeneity, and the large-scale structure of mantle convection: An investigation of an enduring paradox. *Annual Review of Earth and Planetary Sciences* 16: 477–541.
Sizova E, Gerya T, Brown M, and Perchuk LL (2010) Subduction styles in the precambrian: Insight from numerical experiments. *Lithos* 116(3–4): 209–229.
Skemer P, Warren JM, Kelemen PB, and Hirth G (2010) Microstructural and rheological evolution of a mantle shear zone. *Journal of Petrology* 51: 43–53.

Sleep N (1990) Hotspots and mantle plumes: Some phenomenology. *Journal of Geophysical Research* 95: 6715–6736.
Solomatov V (1995) Scaling of temperature dependent and stress dependent viscosity convection. *Physics of Fluids* 7: 266–274.
Sornette D and Pisarenko V (2003) Fractal plate tectonics. *Geophysical Research Letters* 30(3: 1105). http://dx.doi.org/10.1029/2002GL015043.
Sotin C and Labrosse S (1999) Three-dimensional thermal convection in an iso-viscous, infinite Prandtl number fluid heated from within and from below: Applications to the transfer of heat through planetary mantles. *Physics of the Earth and Planetary Interiors* 112(3–4): 171–190. http://www.sciencedirect.com/science/article/pii/S0031920199000047.
Southam G and Westall F (2007) Geology, life and habitability. In: Spohn T and Schubert G (Editor-in-Chief) (eds.) *Treatise on Geophysics*. *Planets and Moons*, vol. 10, pp. 421–437. Amsterdam: Elsevier.
Spiegelman M and McKenzie D (1987) Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth and Planetary Science Letters* 83: 137–152.
Squyres SW, Grotzinger JP, Arvidson RE, et al. (2004) In situ evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* 306(5702): 1709–1714. http://www.sciencemag.org/content/306/5702/1709.abstract.
Stein C, Schmalzl J, and Hansen U (2004) The effect of rheological parameters on plate behaviour in a self-consistent model of mantle convection. *Physics of the Earth and Planetary Interiors* 142: 225–255.
Stein C and Stein S (1992) A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* 359: 123–129.
Steinberger B and O'Connell R (1998) Advection of plumes in mantle flow: Implications for hotspot motion, mantle viscosity and plume distribution. *Geophysical Journal International* 132: 412–434.
Stern RJ (2004) Subduction initiation: Spontaneous and induced. *Earth and Planetary Science Letters* 226: 275–292.
Stern R and Bloomer S (1992) Subduction zone infancy: Examples from the eocene Izu-Bonin-Mariana and Jurassic California arcs. *Geological Society of America Bulletin* 104: 1621–1636.
Strutt JW (Lord Rayleigh) (1916) On convective currents in a horizontal layer of fluid when the higher temperature is on the under side. *Philosophical Magazine* 32: 529–546.
Su W and Dziewonski A (1992) On the scale of mantle heterogeneity. *Physics of the Earth and Planetary Interiors* 74: 29–54.
Tackley P (1996a) Effects of strongly variable viscosity on three-dimensional compressible convection in planetary mantles. *Journal of Geophysical Research* 101: 3311–3332.
Tackley P (1996b) On the ability of phase transitions and viscosity layering to induce long-wavelength heterogeneity in the mantle. *Geophysical Research Letters* 23: 1985–1988.
Tackley P (1998) Self-consistent generation of tectonic plates in three-dimensional mantle convection. *Earth and Planetary Science Letters* 157: 9–22.
Tackley P (2000a) Mantle convection and plate tectonics: Toward and integrated physical and chemical theory. *Science* 288: 2002–2007.
Tackley P (2000b) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 1. Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* 1: 1021. http://dx.doi.org/10.1029/2000GC000036.
Tackley P (2000c) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 2. Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems* 1: 1026. http://dx.doi.org/10.1029/2000GC000043.
Tackley P (2000d) The quest for self-consistent generation of plate tectonics in mantle convection models. In: Richards MA, Gordon R, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*. *Geophysical Monograph Series*, vol. 121, pp. 47–72. Washington, DC: American Geophysical Union.
Tackley P, Ammann M, Brodholt J, Dobson D, and Valencia D (2013) Mantle dynamics in super-earths: Post-perovskite rheology and self-regulation of viscosity. *Icarus* 225(1): 50–61. http://www.sciencedirect.com/science/article/pii/S0019103513001231.
Tackley P, Stevenson D, Glatzmaier GA, and Schubert G (1993) Effects of an endothermic phase transition at 670 km depth in a spherical model of convection in the earth's mantle. *Nature* 361: 699–704.
Tackley P, Stevenson D, Glatzmaier GA, and Schubert G (1994) Effects of a multiple phase transitions in a 3 dimensional spherical model of convection in earth's mantle. *Journal of Geophysical Research* 99: 15877–15901.
Thatcher W and England P (1998) Ductile shear zones beneath strike-slip faults: Implications for the thermomechanics of the San Andreas Fault zone. *Journal of Geophysical Research* 103: 891–905.
Thielmann M and Kaus BJP (2012) Shear heating induced lithospheric-scale localization: Does it result in subduction? *Earth and Planetary Science Letters* 359–360: 1–13.
Thompson B (Count Rumford) (1797) The propagation of heat in fluids. *Bibliotheque Britannique (Science et Arts)* 90: 97–200, Reprinted in Brown, 1967.
Thorkelson DJ (1996) Subduction of diverging plates and the principles of slab window formation. *Tectonophysics* 255(1–2): 47–63. http://www.sciencedirect.com/science/article/pii/0040195195001069.
Torsvik T, Smethurst M, Meert J, et al. (1996) Continental break-up and collision in the Neoproterozoic and Palaeozoic – A tale of Baltica and Laurentia. *Earth Science Reviews* 40(3–4): 229–258. http://www.sciencedirect.com/science/article/pii/0012825296000086.
Toth G and Gurnis M (1998) Dynamics of subduction initiation at preexisting fault zones. *Journal of Geophysical Research* 103: 18053–18067.
Tozer D (1972) The present thermal state of the terrestrial planets. *Physics of the Earth and Planetary Interiors* 6: 182–197.
Tozer D (1985) Heat transfer and planetary evolution. *Geophysical Surveys* 7: 213–246.
Trompert R and Hansen U (1998) Mantle convection simulations with rheologies that generate plate-like behavior. *Nature* 395: 686–689.
Turcotte DL (1993) An episodic hypothesis for Venusian tectonics. *Journal of Geophysical Research, Planets* 98(E9): 17061–17068. http://dx.doi.org/10.1029/93JE01775.
Turcotte D, Morein G, Roberts D, and Malamud B (1999) Catastrophic resurfacing and episodic subduction on Venus. *Icarus* 139(1): 49–54. http://www.sciencedirect.com/science/article/pii/S0019103599960840.
Turcotte D and Morgan JP (1992) The physics of magma migration and mantle flow beneath a mid-ocean ridge. In: Morgan JP, Blackman D, and Sinton J (eds.) *Mantle Flow and Melt-Generation at Mid-Ocean Ridges*. *Geophysical Monograph Series*, vol. 71, pp. 155–182. Washington, DC: American Geophysical Union.
Turcotte D and Oxburgh E (1967) Finite amplitude convection cells and continental drift. *Journal of Fluid Mechanics* 28: 29–42.
Turcotte D and Oxburgh E (1972) Mantle convection and the new global tectonics. *Annual Review of Fluid Mechanics* 4: 33–66.
Turcotte D and Schubert G (2002) *Geodynamics*. Cambridge: Cambridge University Press.
Urai J, Means W, and Lister G (1986) Dynamic recrystallization in minerals. In: Hobbs B and Heard H (eds.) *Mineral and Rock Deformation: Laboratory Studies*, pp. 166–199. Washington, DC: American Geophysical Union.
Valencia D and O'Connell RJ (2009) Convection scaling and subduction on earth and super-earths. *Earth and Planetary Science Letters* 286(3–4): 492–502.
Valencia D, O'Connell R, and Sasselov D (2007) Inevitability of plate tectonics on super-earths. *Astrophysics Journal* 670: L45–L48.
Valley JW, Peck WH, King EM, and Wilde SA (2002) A cool early earth. *Geology* 30(4): 351–354.http://geology.gsapubs.org/content/30/4/351.abstract.
van der Hilst R, Engdahl R, Spakman W, and Nolet G (1991) Tomographic imaging of subducted lithosphere below northwest pacific island arcs. *Nature* 353: 37–43.
van der Hilst R, Widiyantoro S, and Engdahl E (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386: 578–584.
van Heck H and Tackley P (2008) Planforms of self-consistently generated plates in 3D spherical geometry. *Geophysical Research Letters* 35: L19312. http://dx.doi.org/10.1029/2008GL035190.
van Heck H and Tackley P (2011) Plate tectonics on super-earths: Equally or more likely than on earth. *Earth and Planetary Science Letters* 310(3–4): 252–261.
van Hunen J and Moyen J-F (2012) Archean subduction: Fact or fiction? *Annual Review of Earth and Planetary Sciences* 40(1): 195–219. http://www.annualreviews.org/doi/abs/10.1146/annurev-earth-042711-105255.
van Hunen J and van den Berg AP (2008) Plate tectonics on the early earth: Limitations imposed by strength and buoyancy of subducted lithosphere. *Lithos* 103(1–2): 217–235.
Vigny C, Ricard Y, and Froidevaux C (1991) The driving mechanism of plate tectonics. *Tectonophysics* 187: 345–360.
Vine F and Matthews D (1963) Magnetic anomalies over oceanic ridges. *Nature* 199: 947–949.
Walker J, Hayes P, and Kasting J (1981) A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86: 9776–9782.
Warren JM and Hirth G (2006) Grain size sensitive deformation mechanisms in naturally deformed peridotites. *Earth and Planetary Science Letters* 248(1–2): 438–450.
Watts A, Karner G, and Steckler M (1982) Lithosphere flexure and the evolution of sedimentary basins. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences* 305: 249–281.

Weertman J and Weertman J (1975) High temperature creep of rock and mantle viscosity. *Annual Review of Earth and Planetary Sciences* 3: 293–315.
Wegener A (1924) *The Origin of Continents and Oceans*. London: Methuen and Co.
Weinstein S (1993) Catastrophic overturn of the earth's mantle driven by multiple phase changes and internal heating. *Geophysical Research Letters* 20: 101–104.
Weinstein S (1996) Thermal convection in a cylindrical annulus with a non-Newtonian outer surface. *PAGEOPH* 146: 551–572.
Weinstein S (1998) The effect of convection planform on the toroidal–poloidal energy ratio. *Earth and Planetary Science Letters* 155: 87–95.
Weinstein S and Christensen U (1991) Convection planforms in a fluid with a temperature-dependent viscosity beneath a stress-free upper boundary. *Geophysical Research Letters* 18: 2035–2038.
Weinstein S and Olson P (1989) The proximity of hotspots to convergent and divergent plate boundaries. *Geophysical Research Letters* 16: 433–436.
Weinstein S and Olson P (1990) Planforms in thermal convection with internal heat sources at large Rayleigh and Prandtl numbers. *Geophysical Research Letters* 17: 239–242.
Weinstein S and Olson P (1992) Thermal convection with non-Newtonian plates. *Geophysical Journal International* 111: 515–530.
Wessel P and Kroenke LW (2008) Pacific absolute plate motion since 145 Ma: An assessment of the fixed hot spot hypothesis. *Journal of Geophysical Research - Solid Earth* 113(B6). http://dx.doi.org/10.1029/2007JB005499.
White D (1988) The planforms and onset of convection with temperature-dependent viscosity. *Journal of Fluid Mechanics* 191: 247–286.
White S, Burrows S, Carreras J, Shaw N, and Humphreys F (1980) On mylonites in ductile shear zones. *Journal of Structural Geology* 2: 175–187.
Whitehead J (1976) Convection models: Laboratory versus mantle. *Tectonophysics* 35: 215–228.
Whitehead J and Gans R (1974) A new, theoretically tractable earthquake model. *Geophysical Journal of the Royal Astronomical Society* 39: 11–28.
Whitehead J and Parsons B (1978) Observations of convection at Rayleigh numbers up to 760,000 in a fluid with a large Prandtl number. *Geophysical and Astrophysical Fluid Dynamics* 19: 201–217.
Wilson J (1966) Did the Atlantic close and then re-open? *Nature* 211: 676–681.
Wolfe C, Bjarnason I, VanDecar J, and Solomon S (1997) Seismic structure of the iceland mantle plume. *Nature* 385: 245–247.
Wolfe C, Solomon S, Laske G, et al. (2009) Mantle shear-wave velocity structure beneath the Hawaiian hotspot. *Science* 326: 1388–1390.
Yin A (2012) Structural analysis of the Valles Marineris fault zone: Possible evidence for large-scale strike-slip faulting on mars. *Lithosphere* 4(4): 286–330.
Yoshida M (2013) Mantle temperature under drifting deformable continents during the supercontinent cycle. *Geophysical Research Letters* 40(4): 681–686.
Yuen D, Fleitout L, Schubert G, and Froidevaux C (1978) Shear deformation zones along major transform faults and subducting slabs. *Geophysical Journal of the Royal Astronomical Society* 54(1): 93–119.
Zhang S and Yuen D (1995) Intense local toroidal motion generated by variable viscosity compressible convection in 3-D spherical shell. *Geophysical Research Letters* 23: 3135–3138.
Zhong S and Gurnis M (1993) Dynamic feedback between a continent-like raft and thermal-convection. *Journal of Geophysical Research – Solid Earth* 98(B7): 12219–12232.
Zhong S and Gurnis M (1995a) Mantle convection with plates and mobile, faulted plate margins. *Science* 267: 838–843.
Zhong S and Gurnis M (1995b) Towards a realistic simulation of plate margins in mantle convection? *Geophysical Research Letters* 22: 981–984.
Zhong S and Gurnis M (1996) Interaction of weak faults and non-Newtonian rheology produces plate tectonics in a 3d model of mantle flow. *Nature* 383: 245–247.
Zhong S, Gurnis M, and Moresi L (1998) Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models. *Journal of Geophysical Research* 103: 15255–15268.
Zhong S, Zuber M, Moresi L, and Gurnis M (2000) Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *Journal of Geophysical Research* 105: 11063–11082.
Zindler A and Hart S (1986) Chemical geodynamics. *Annual Review of Earth and Planetary Sciences* 14: 493–571.

7.08 The Dynamics and Convective Evolution of the Upper Mantle

EM Parmentier, Brown University, Providence, RI, USA

© 2015 Elsevier B.V. All rights reserved.

7.08.1	**Introduction**	319
7.08.2	**Cooling the Mantle from Above**	319
7.08.2.1	An Historical Perspective	319
7.08.2.2	Convective Instability Versus Heating by Hot Spots: Depth–Age and Heat Flow	320
7.08.2.3	Geoid Height as a Measure of Upper Mantle Thermal Structure	322
7.08.2.3.1	Intraplate seismicity and plate elasticity as an indicator of thermal structure	323
7.08.2.4	Seismic Velocity and Attenuation as Measures of Upper Mantle Structure and Flow	323
7.08.2.5	Upper Mantle Electrical Conductivity: Water Content and Temperature	326
7.08.2.6	The Age of Thermal Convective Instability Beneath Oceanic Lithosphere	326
7.08.2.7	Implications from Recent Theoretical and Experimental Studies of Convective Instability	327
7.08.3	**Convective Instability and Melting in the Upper Mantle**	329
7.08.3.1	Intraplate Volcanism as an Indicator of Upper Mantle Convective Activity	329
7.08.3.2	Other Possible Mechanisms for Intraplate Volcanism	330
7.08.3.2.1	Melting in convectively driven upwellings	330
7.08.3.2.2	Melting due to shear-driven upwelling	330
7.08.3.3	Buoyant Decompression Melting	330
7.08.4	**Upwelling and Melting Beneath Oceanic Spreading Centers**	331
7.08.4.1	Melting, Melt Extraction, and the Chemical Lithosphere	331
7.08.4.2	Influence of Melt Extraction Mechanism	331
7.08.4.3	2D Versus 3D Upwelling and the Spreading Rate Dependence of Seafloor Structure	332
7.08.5	**Summary**	332
References		334

7.08.1 Introduction

The thermal structure of the oceanic upper mantle plays an important role in the larger-scale dynamics of the mantle. Fortunately, this region of the Earth is one that is accessible to observation by a variety of geophysical tools and so can provide a basis for understanding solid-state convective instability, with implications for even more general settings than the Earth's upper mantle. It can be argued that the Earth's upper mantle is the best natural laboratory available to us to study the convective evolution of planetary interiors.

The goal of this chapter is to summarize current understanding of the evolution of the upper mantle based on observations and theoretical predictions, beginning with some of the earliest observations on how cooling the mantle from above would be expressed in seafloor depth, heat flow, and geoid height as functions of age. It is not possible to consider the role of convective instability in the upper mantle in isolation from other forms of larger-scale mantle flow. While no general agreement has yet emerged on how various scales of convection may affect the upper mantle, small-scale thermal convective instability generated within the upper mantle provides a viable explanation for many important observations. Convection in the upper mantle may be driven by buoyancy resulting both from cooling at the surface and by melting. Mantle cooling is expressed in oceanic heat flow and in geoid height and seafloor depth through the effect of cooling on mantle density.

Buoyancy due to melting may be expressed in intraplate volcanism. Only a few long-lived, age-progressive, linear volcanic chains meet all the criteria of the deep mantle plume-fixed hot spot hypothesis (see 00010, this volume). Decompression melting in small-scale thermally driven convective upwellings has been one favored explanation for volcanic ridges aligned with plate motion. Seafloor volcanism may also reflect spontaneously generated instability driven by decompression melting, which has been termed 'magmatic convective storms.' Buoyant decompression melting beneath moving plates is one possible mechanism for the abundant short-lived island chains and volcanic ridges identified on the Earth's seafloor.

Finally, the role of mantle upwelling and melting beneath spreading centers is important to consider, in part because it is responsible for the generation of the oceanic crust and is a factor controlling spreading center structure. However, melting and melt extraction under spreading centers leave behind residual mantle that is transported away from the spreading center. Compositional changes are expected to affect viscosity or creep rate and create a stable chemical stratification, thus fundamentally influencing convective instability everywhere in the upper mantle (Kohlstedt, Vol. 2).

7.08.2 Cooling the Mantle from Above

7.08.2.1 An Historical Perspective

Turcotte and Oxburgh (1967) identified the moving lithospheric plates (Wessel, Vol. 6) as the conductive thermal boundary layer at the top of mantle convection cells, thus

 http://dx.doi.org/10.1016/B978-0-444-53802-4.00131-7

establishing a physical relationship between upper mantle thermal structure and mantle dynamics. During the emergence of plate tectonics, measurements of seafloor depth and heat flow in the oceans provided the first direct evidence on the thermal structure of the upper mantle. Heat flow generally decreased with crustal age but showed great variability, particularly at young ages. This variability is now understood to be due primarily to hydrothermal circulation of seawater through permeable crustal rocks. In contrast, seafloor depths showed a much more systemic variation with seafloor age. The age dependence of isostatic seafloor depth, which depends on the depth-averaged temperature, was recognized as a stronger observational constraint than heat flow on upper mantle thermal structure (Langseth et al., 1966; McKenzie, 1967; McKenzie and Sclater, 1969; Sleep, 1969; Vogt and Ostenso, 1967).

The thermal structure that develops due to conductive cooling should be, at least to first order, only a function of crustal age, rather than an independent function of both distance from spreading center and plate velocity. Observations of seafloor depth and age generally confirmed this age dependence. The Turcotte and Oxburgh boundary layer hypothesis, in the simple form suggested originally, indicated that the thermal boundary layer should continue to thicken as the square root of age so that old seafloor would continue to subside. This did not appear to explain observations showing a relatively uniform depth of old seafloor that required that the thermal boundary layer evolve to a nearly constant thickness. Models for upper mantle thermal structure that were consistent with average depth of seafloor as a function of age from the growing collection of observations treated the conductive cooling of horizontally moving upper mantle, as in the simple boundary layer theory, but with the added assumption of a uniform temperature at a prescribed depth (Langseth et al., 1966; McKenzie, 1967). A relatively uniform seafloor depth at old ages requires a mechanism to transport heat to the bottom of the thermal boundary layer, thus reducing the rate at which it thickens. This set the stage for research in the following several decades and is still a source of continuing study and debate.

7.08.2.2 Convective Instability Versus Heating by Hot Spots: Depth–Age and Heat Flow

Both small-scale convective instability of the cool thermal boundary and heating at hot spots (McNutt, Vol. 1; 00009, this volume) may play a role in transporting heat to the bottom of the thermal boundary layer, but no general consensus has yet emerged on a single mechanism of heating that can explain all available observations. Ideally, observed variations in seafloor depth might constrain the relative importance of these two mechanisms, but such interpretations are limited by the relatively small amount of old seafloor with a well-determined age, by the uncertainty in correcting for the thickness and load of sediments, and by unidentified crustal thickness variations. Many earlier studies have shown that simply averaging all depth measurements at a given age indicates that old seafloor is, on average, shallower than would be predicted by conductive cooling alone. Several recent studies have continued to refine earlier interpretations of a seafloor depth–age relationship (Crosby and McKenzie, 2009; Crosby et al., 2006; Hillier and Watts, 2005). As shown in **Figure 1** and as further discussed in the succeeding text, differing methodology in these

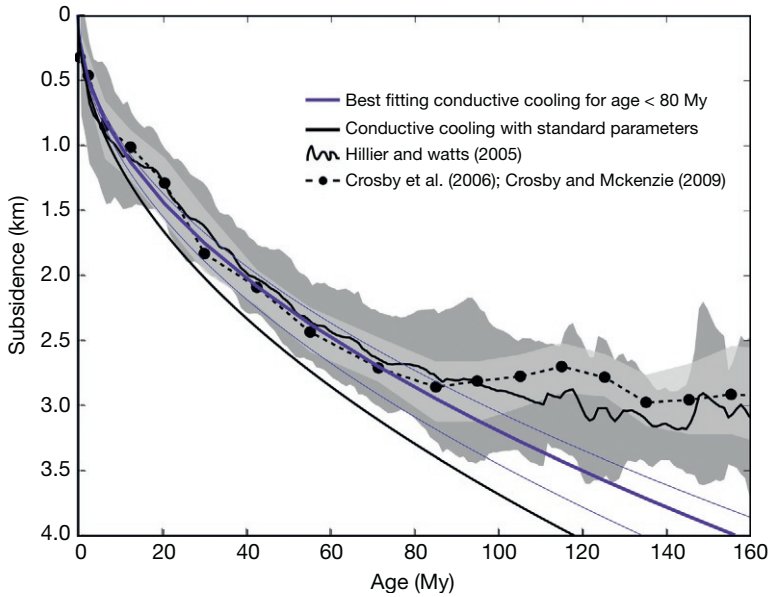


Figure 1 Estimates of seafloor depth as a function of age for various oceans, based on depths corrected for sediment and crustal thickness variations, compared to conductive cooling models. The black curve is from Hillier and Watts (2005) with the darker shaded region showing one standard deviation about the mean value at each age. The black dashed curve is from Crosby et al. with the light shaded region showing the variation in topography observed in a given age bin. Different criteria used to define 'normal' seafloor, discussed in the text, account differences shown. The solid black line corresponds to conductive cooling with nominal values of thermal diffusivity (10^{-6} m^2 s^{-1}), thermal expansivity (3×10^{-6} K^{-1}), mantle density (3300 kg m^{-3}), and mantle temperature (1350 °C) giving a subsidence rate of 386 m Ma$^{-1/2}$. The blue curves correspond to a best-fitting subsidence rate (325 ± 20 m Ma$^{-1/2}$) for seafloor younger than 70 My.

studies leads to different depth–age estimates. However, in both studies, old seafloor appears to be at least several hundred meters shallower than the depth predicted by a purely conductive cooling model that fits well at ages <70 My. So the long-standing question is: why is old seafloor often shallower than predicted by conductive cooling alone, and how is it possible to infer the physical process(es) responsible?

Heating due to hot spots should correlate with the length of time that a given area of seafloor has spent in the proximity of a hot spot (Crough, 1975). Heestand and Crough (1981) sorted the depth of seafloor in the North Atlantic by distance from the nearest hot spot track. Comparing depth–age for seafloor in constant distance ranges indicated no flattening at old ages. Hayes (1988) noted that in the South Atlantic, where hot spot influences appear to be less significant, depth followed a purely conductive cooling curve to ages approaching 120 My but noted the complication of a persistent asymmetry in apparent subsidence rate across the spreading axis. An asymmetry in the apparent subsidence rate was also found in the southeast Indian Ocean (Hayes, 1988) and across the East Pacific Rise (Eberle and Forsyth, 1995). One possibility may be that this asymmetry reflects the dynamic topography of larger-scale mantle flow on which the depth–age subsidence is superimposed (Jian Lin, Vol. 6).

The Pacific Plate contains not only relatively large areas of old seafloor but also numerous seamounts and hot spot tracks (e.g., Wessel and Lyons, 1997). Studies like those cited in **Figure 1** have compiled seafloor depth–age curves for the Pacific and other ocean basins, often with differing results depending on the criteria used to exclude seafloor affected by volcanism or other processes. Schroeder (1984) found that eliminating data within 800 km of hot spot tracks resulted in a good correlation of depth with square root of age for ages <80 My. All seafloor older than 80 My was shallower than predicted by conductive cooling alone but was within 800 km of hot spots or hot spot tracks. Renkin and Sclater (1988) analyzed the effect of uncertainties in basement age and sediment thickness on depth–age correlations in the North Pacific. Arguing that volcanic constructs always result in shallower seafloor, they proposed that modal seafloor depth, or more precisely a range enclosing 2/3 of the measured depths at any age, provides an estimate of subsidence associated with cooling least biased by volcanism. Modal depths increase with age more slowly than predicted by conductive cooling for ages exceeding about 80 My. Plotting modal depths with age along a corridor of the Pacific seafloor containing several swell-like features, Renkin and Sclater argued that not even the deepest depths fall along a conductive cooling model. The studies of Hillier and Watts (2005) and Crosby et al. (2005) shown in **Figure 1** are among the more recent studies illustrating that differing criteria for identifying normal seafloor (not affected by processes other than simple aging) lead to different estimates of a depth–age relationship. Hillier and Watts (2005) employed a filtering technique that appears to be an automation of visual inspection. After excluding by eye seafloor thought to be affected by hot spot chains and oceanic plateaux, Crosby et al. used the absence of gravity anomalies to identify seafloor included in their depth–age compilation seafloor. Korenaga and Korenaga (2008) used both distance from hot spot chains and statistical correlation criteria. Both Crosby et al. (2005) and Korenaga and Korenaga (2008) found best-fitting square root of age subsidence rate that is smaller than predictions based on mineral physics data. Korenaga and Korenaga (2008) attributed this to a lower effective thermal expansivity of a viscoelastic mantle (Korenaga, 2007). An interesting conclusion of both Crosby et al. (2005) and Korenaga and Korenaga (2008) is a strong correlation between zero-age depth and subsidence rate. The possible cause or implications of this correlation have not yet been addressed.

Small-scale convective instability of the thermal boundary layer, in addition to hot spots, may advectively transport heat to the bottom of the conductive boundary layer. The possible importance of small-scale convection was first suggested on the basis of averaged depth–age curves that deviate from purely conductive cooling at a seafloor age of about 70–80 My, which was presumed to correspond to the onset of convective instability (Parsons and McKenzie, 1978). Beyond this age, convective heat transfer was thought to maintain the nearly constant thermal boundary layer thickness implied by the plate model (Davis and Lister, 1974; Parsons and Sclater, 1977; Stein and Stein, 1992). Plate thickness must be consistent with the heat flow measured on old seafloor (Davis et al., 1984; Lister et al., 1990; Nagihara et al., 1996) and its depth. **Figure 2** from Nagihara et al. (1996) shows heat flow values for old seafloor in the northwestern Pacific and in the western North Atlantic, both areas where high-quality heat flow and data needed to correct depth for sediment loading and crustal thickness variations are available. Also shown are predictions from two versions of the plate model and purely conductive cooling, all of which are thought to provide reasonable fits to the depth–age relationship of younger seafloor. Heat flow is higher and seafloor is consistently shallower than predicted by conductive cooling. The plate model of parameters of Parsons and Sclater (1977) with a plate thickness 125 km and a mantle temperature 1350 °C fits the depths well but underestimates the heat flow. In contrast, the hotter and thinner plate model of Stein and Stein (1992) with a plate thickness 95 km and a mantle temperature 1450 °C fits the heat flow well but underestimates old seafloor depth. It is worth noting that this mantle temperature is significantly higher than that estimated from melting in adiabatically upwelling mantle beneath spreading centers, frequently inferred to be 1325–1350 °C. In contrast to thermal evolutions with a constant temperature at a prescribed depth, constant heat flux from below (Doin and Fleitout, 1996) also appears to fit the observations within some appropriately defined uncertainties. Goutorbe and Hillier (2013) provided a recent comparison of the fit of these two models to observations.

Thus, numerous studies have found that seafloor older than 70–90 My reaches depths several hundred meters shallower than would be predicted by simple conductive cooling of the upper mantle; and this has typically been explained by some form of heat transfer from the deeper mantle. The usual assumption of a constant crustal thickness implies melting due to upwelling of mantle beneath spreading centers with a time-independent potential temperature and composition. For plausible ranges of crustal and mantle densities, only an additional 400 m of crust would be required to elevate the seafloor by 100 m. Given the small number of accurate measurements of crustal thickness in old oceanic areas particularly the large

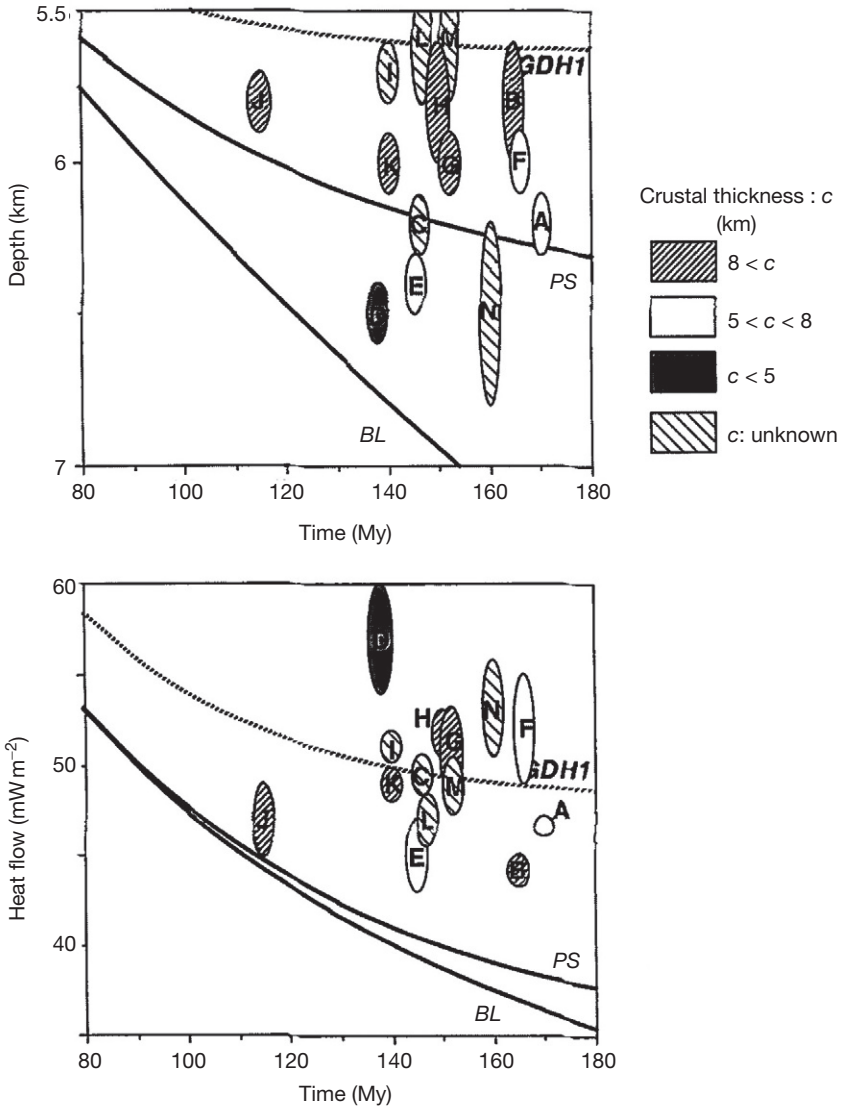


Figure 2 Seafloor basement depth (top), corrected for sediment load and crustal thickness where available, and heat flow (bottom) at sites in the western North Pacific (A–X) and northwestern Atlantic (X–Z) as functions of age from Nagihara et al. (1996). The location of specific sites is identified in the Table 1 and Figure 1 of Nagihara et al. Model curves for the plate models (PS and GDH1) and half-space cooling (BL).

areas of old seafloor in the western Pacific (Oikawa et al., 2010; White et al., 1992), the possibility of a secular increase in crustal thickness and associated depleted mantle lithosphere due to a change in potential temperature or composition of the upper mantle might deserve more consideration. The large number of volcanic seamounts in parts of the Pacific Plate older than 80 Ma has been long recognized and was recently discussed by Hillier (2006). Conrad et al. (2011) attributed this to a time of rapid plate motion and mantle shearing. Alternatively, this has been attributed to a global-scale change in upper mantle thermal structure (Larson, 1991) that could explain a range of geologic observations (Larson and Erba, 1999). Additional attention to measures of upper mantle thermal and compositional structure in addition to depth–age will be required to distinguish between various explanations of seafloor depth evolution.

7.08.2.3 Geoid Height as a Measure of Upper Mantle Thermal Structure

As discussed previously, the earliest interpretations of the thermal evolution of the upper mantle relied largely on heat flow and isostatic seafloor depth, with the isostatic assumption justified by the absence of free-air gravity anomalies correlated with the depth–age variation. Following the advent of sea surface elevation measurements from satellite altimetry, the long-wavelength geoid (Jekeli, Vol. 3) and gravity provided further constraint on upper mantle thermal structure (Chase, 1985; Smith, 1998). Isostatic seafloor depth measures the density averaged over depth in the mantle; geoid height measures the first moment of the density distribution with depth (Haxby and Turcotte, 1978). Conductive cooling predicts that geoid height decreases linearly with age

(Haxby and Turcotte, 1978) as they observed along a GEOS-3 altimetry track in the North Atlantic. Also, using early GEOS-3 altimetry data, Sandwell and Schubert (1980) found that geoid height decreases approximately linearly with seafloor age for ages less than about 80 My in the Atlantic and southeast Indian spreading centers with geoid slopes in the range −0.131 to −0.149 m My^{-1}, generally consistent with a conductive cooling thermal boundary layer. In contrast, the geoid height predicted by the plate model should flatten over old seafloor. Sandwell and Schubert (1980) found that geoid height in the southeast Pacific was not consistent with a simple linear geoid–age relationship but decreased rapidly with age. The decrease was much more rapid than predicted by the plate model used to explain flattening of old seafloor. This is further discussed in the succeeding text.

The geoid height is sensitive to density variations over much larger spatial scales than gravity anomalies. Therefore, density variations throughout the Earth, not just in the upper mantle, contribute to geoid height variations. The difficulty in using geoid as a constraint on the age dependence of the upper mantle is to separate variations in geoid height with seafloor age from other contributions, for example, from oceanic swells or hot spots. One approach to filtering out the long-wavelength contributions is to measure the change in geoid height across fracture zones that juxtapose seafloor of different ages (Crough, 1979a; Detrick, 1981). If geoid height decreases linearly with age as expected for conductive cooling, the change in geoid height across a fracture zone of constant age offset should remain constant as seafloor on each side of the fracture zone ages. In contrast, the plate model should show a change in geoid height across the fracture zone that decreases with age and that vanishes once thermal boundary layer thickness becomes constant. Observations suggest that geoid slope decreases with age, but it appears to do so at a much earlier age than that at which the seafloor flattens. Studies of variations of geoid height with age include Cazenave (1984), Driscoll and Parsons (1988), and Freedman and Parsons (1990). Geoid height as a function of age from Cazenave (1984), plotted as geoid–age slope, is shown in **Figure 3**.

Determining the geoid height change with age using measurements across fracture zones is elegant in concept but difficult in practice. Mechanical behaviors of lithosphere, both plate flexure associated with differential subsidence (Sandwell and Schubert, 1982a,b) and thermal bending (Parmentier and Haxby, 1986; Wessel and Haxby, 1990), create uncompensated seafloor topography over distances of 50–100 km adjacent to fracture zones. The resulting geoid anomalies may obscure the identification of the change in isostatic geoid height. Despite these uncertainties, geoid height variation with age is not explained by either the half-space cooling or the plate model mantle thermal structure.

7.08.2.3.1 Intraplate seismicity and plate elasticity as an indicator of thermal structure

The age dependence of the apparent elastic thickness of the oceanic lithosphere can also constrain the variation of mantle thermal structure with age. Focal depths of oceanic intraplate earthquakes increase with age following a 650–700 °C isotherm in a conductive cooling model (Bergman and Solomon, 1984; Weins and Stein, 1984). However, most of the seismicity reported occurs at ages <70 My and so does not constrain thermal structure at older ages. Flexure of lithosphere at trenches should also reflect the mechanical thickness. Levitt and Sandwell (1995) indicated that plate thickness increases with age of subducting lithosphere and that the plate in the Stein and Stein (1992) thermal model may be too hot and thin. But the data do not resolve the difference between a simple conductive cooling and a plate model with a thicker plate.

7.08.2.4 Seismic Velocity and Attenuation as Measures of Upper Mantle Structure and Flow

Seismic velocities and attenuation in the upper mantle (Montagner, Vol. 1 and (Rainer) Kind, Vol. 1) also constrain thermal structure and its age evolution. Surface wave dispersion has been used to infer the velocity variation with age and depth (Maggi et al., 2006a,b,c; Nettles, 2006; Nettles and Dziewonski, 2008; Nishimura and Forysth, 1989; Priestly and McKenzie, 2006; Ritzwoller et al., 2004; Weeraratne et al., 2007; Zhou et al., 2006). As shown in **Figure 4**, seismic shear wave velocities generally decrease with depth at a given age and increase with age at a given depth. Seismic velocities, which increase with decreasing temperature, continue to change with age at depths exceeding 150 km. Using the parameterizations of Priestley and McKenzie (2006) as a first attempt to assess the

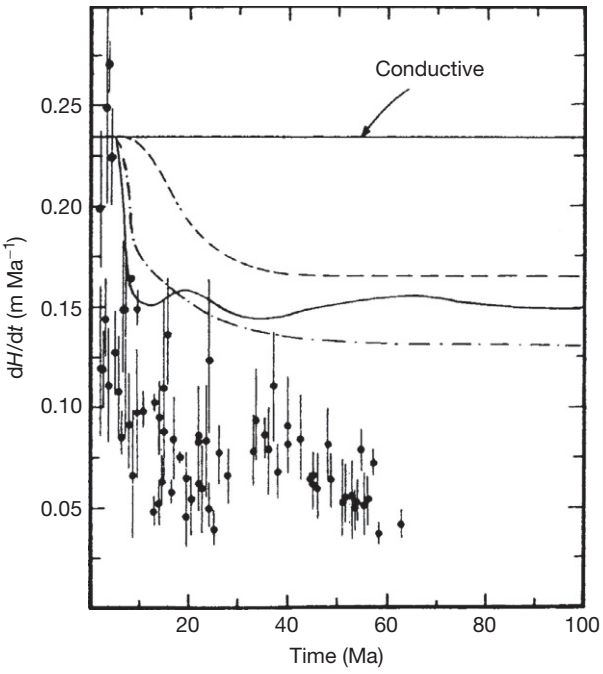


Figure 3 Geoid–age slope as a function of age with data from Cazenave (1984). Pure conductive cooling leads to the constant geoid–age slope as shown. The observations indicate a rapid decline with age. Curves show model predictions from Buck and Parmentier (1986) for convective cooling with an activation energy of 420 kJ mol^{-1} and an activation volume of 10 cm^3 mol^{-1}. Dashed and long–short dashed curves are for a model reference viscosity of 10^{18} Pa s and 5×10^{18} Pa s, respectively. The solid curve is also for a 10^{18} Pa s viscosity but in a wider and more finely discretized model domain.

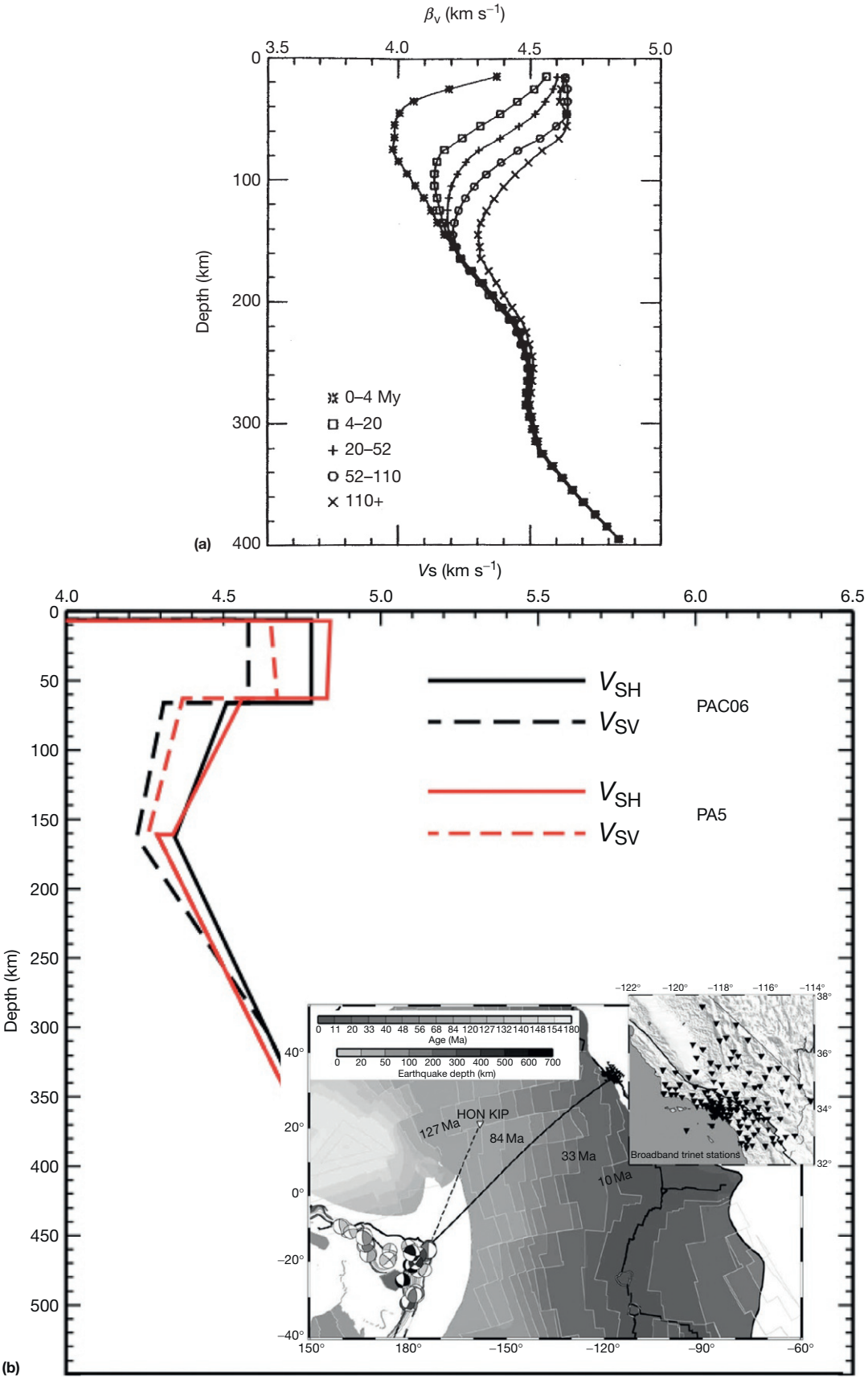


Figure 4 (a) Vertically polarized shear wave velocity in the Pacific as a function of depth in various age intervals determined from Rayleigh wave dispersion (Nishimura and Forsyth, 1989). Note that shear wave velocity changes with age at depths greater than would be consistent with a plate model that fits the depth–age relationship ($\leq$125 km). Changes with age continue beneath seafloor with ages exceeding 110 My. (b) Shear wave velocity as a function of depth derived from *S*-wave reverberations along paths in the Pacific Plate modified from Tan and Helmberger (2007). Velocity is shown for both horizontally (SH) and vertically (SV) polarized shear waves. PA5 is for the corridor from Tonga–Fiji to Hawaii (Gaherty et al., 1996) and PAC06 is for the corridor from Tonga–Fiji to the TriNet array in Southern California (Tan and Helmberger, 2007).

role of temperature alone on seismic velocity, the velocity change of 0.1 km s^{-1} at 150 km depth between 52 and 110 My and >110 My would require cooling in the range of 100–200 °C. Unless some mechanism other than cooling causes the velocity increase, this clearly contradicts the plate model that implicitly assumes that mantle temperature at depths greater than the plate thickness does not change with age. Seismic velocity distributions like those in **Figure 4** may also constrain the depth to which convective instability has cooled or in some other way affected the mantle. Early studies like that of Nishimura and Forsyth (1989) probably do not resolve variations deeper than ~200 km. More recent studies using longer-period (and consequently longer-wavelength) surface waves (Maggi et al., 2006a,b,c; Nettles and Dziewonski, 2008; Zhou et al., 2006), which should resolve velocity variations to greater depth, indicate variations deeper than 200 km. However, the extent to which these variations correlate with plate age has not been established.

Surface waves, because of their long wavelength, have limited depth resolution in sampling mantle velocities. High-frequency body waves are required to detect short-wavelength variations in seismic velocity. Since the pressure, or depth, of phase changes in the upper mantle (Oganov, Vol. 2) depends on temperature, temperature variations should be reflected in variations in the thickness of the mantle transition zone. Using travel times of converted phases and SS precursors, the results of Lawrence and Shearer (2006) (**Figure 6(b)**) show systematic westward thickening of the transition zone in the Pacific, correlating with increasing plate age. This could imply temperature variations associated with convective instabilities beneath cooling plates may extend to depths of at least the top of the transition zone, ~440 km depth.

The G discontinuity is an important feature of the oceanic upper mantle seismic velocity structure that has been detected using reflected and converted body waves. A recent review is given by Rychert et al. (2012). Early work of Gaherty et al. (1996, 1999) and Tan and Helmberger (2007) examined propagation paths in the Pacific shown in **Figure 4(b)**. **Figure 4(b)** also shows shear wave velocity profiles inferred from the analysis of waveforms for propagation along these paths. The G discontinuity at a depth of about 60 km is a prominent feature of the modeled variation of shear wave velocity with depth, defining the top of the low-velocity asthenosphere. This feature has thus often been thought to coincide with the lithosphere–asthenosphere boundary. Using a combination of ScS, multiple S bounces, and surface waves (Gaherty et al., 1999; Tan and Helmberger, 2007), this G discontinuity appears to be due to a sharp (<30 km thick) 6% velocity decrease at relatively constant depth (60 ± 20 km) across a large transect of the central Pacific. More recently, a number of seismic tools including receiver functions, SS precursors, and velocity inversions using waveform modeling of multiple S reflections have also detected a shallow upper mantle seismic discontinuity in the Pacific Ocean. However, these methods give different results in both magnitude and thickness of the 'discontinuity' and where it is detected. High-frequency SS precursors (Schmerr, 2012) detect a sharp (<20 km thick) velocity decrease (≥5%) showing a weak age dependence (~55–75 km) primarily beneath hot spot regions but not elsewhere. ScS reverberations detect a sharp (<30 km thick) velocity decrease (5–14%) at 72–112 km depth beneath a large swaths of the western Pacific but not in the eastern Pacific (Bagley and Revenaugh, 2008). Receiver function imaging of subducting oceanic lithosphere reveals a sharp (<15 km thick) discontinuity (7–8%) (Kawakatsu et al., 2009) in the 44–85 km depth range (Kawakatsu et al., 2009; Kumar and Kawakatsu, 2011), which is imaged on the old Pacific and younger Philippine Sea Plate. A discontinuity at 25–130 km depth is imaged beneath large portions of the Pacific, increasing in depth with crustal age (Rychert and Shearer, 2011). A sharp discontinuity is detected beneath young seafloor, but not necessarily everywhere. One possible explanation for these variable results may be the presence of anisotropic seismic velocity (Rychert et al., 2012). Previous methods applied considering waves propagating at different azimuths show propagation direction variation in the inferred velocity structure. Anisotropy may thus be an important factor in imaging and constraining discontinuity structure in the oceanic upper mantle. It will be important to consult recent publications in the rapidly developing area.

Seismic anisotropy (Mainprice, Vol. 2; Park Vol. 1) in the upper mantle also provides a possible indication of mantle flow and could have important implications for understanding convective instability. Azimuthal anisotropy has been long recognized (cf. Montagner and Tanimoto, 1991; Nishimura and Forsyth, 1989), but its depth variation and the relative contributions from the lithosphere and asthenosphere are not yet resolved. In a simple unidirectional parallel flow, the seismic fast direction is usually assumed to lie close to the shear plane and in the flow direction. In more complex flows, the seismically fast direction has been assumed to coincide with the local direction of maximum accumulated elongation. However, the strength and local direction of anisotropy will be determined by the rates at which preferred orientation of mineral grains is created and destroyed by shear and dynamic recrystallization, respectively (Chastel et al., 1993; Kaminski and Ribe, 2001, 2002; Wenk and Tomé, 1999). Several recent global tomography studies have also reported radial anisotropy that is more pronounced beneath the Pacific Plate than elsewhere in the oceanic upper mantle (e.g., Ekstrom and Dziewonski, 1998; Gaboret et al., 2003; Nettles and Dziewonski, 2008). Could this radial anisotropy be a consequence of seismically fast directions vertically aligned in the upwellings and downwellings of convective motion? Hopefully, questions of this type can be resolved by continuing study.

The origin of the seismically defined low-velocity, high-attenuation asthenosphere is not fully resolved (Stixrude, Vol. 2). The possible effects of small amounts of melt (Boehler, Vol. 2) have been frequently discussed, but recent studies indicate that the presence of melt may not be required (Priestly and McKenzie, 2006; Stixrude and Lithgow-Bertelloni, 2006a,b) based on the new understanding of factors controlling shear wave velocities at seismic frequencies now emerging. Stress relaxation at grain boundaries is thought to be an important anharmonic effect that introduces a frequency–grain size variation and a temperature dependence to the elastic moduli in addition to that of high-frequency (harmonic) variations with temperature and pressure (Cooper, 2002; Faul and Jackson, 2005; Gribb and Cooper, 1998; Jackson et al., 2002). Variations in grain size and water content could thus contribute to the velocity variation with depth like those shown in **Figure 4**.

Explaining the origin of the G discontinuity has been a recent focus of interest. The presence of this seismic

discontinuity could have important implications for the origin of the asthenosphere. While seismic velocity decreases with increasing temperature, temperature gradients at the top of the asthenosphere are generally expected to be too small to explain the small width of the G discontinuity. Three different mechanisms have been discussed for the origin of the G discontinuity and perhaps the asthenosphere beneath it: (a) variations of seismic velocity with grain size and water content, (b) the development or changing directions of seismic anisotropy with depth, and (c) the presence of small amounts of melt.

Increased seismic attenuation in the asthenosphere may account for the reduced seismic velocity there, as discussed earlier. But whether attenuation can explain the rapid variation of velocity with depth required to create the G discontinuity is less clear. Karato (2012) provided a recent review of possible mineral physics mechanisms that may be responsible for seismic velocity variations. Seismic velocity is controlled by anharmonic (elastic) and anelastic mechanisms. The dependence on temperature and pressure for the anharmonic component is well known. Anelastic mechanisms are due to frequency-dependent attenuation that lowers seismic velocity. At high frequency, anelastic relaxation by elastically accommodated grain boundary sliding is characterized by a sharp peak at a particular frequency. At lower frequency, a diffuse absorption band is caused by diffusional accommodation of elastic stresses by grain boundary sliding. The frequency of the absorption peak is expected to increase with water content and temperature, although this has not yet been fully verified by experiments. The hypothesis proposed by Karato (2012) for the origin of the G discontinuity is then: in the cool and dry lithosphere, the absorption peak occurs at a frequency exceeding that of seismic body waves so that the shear modulus is unrelaxed (corresponding to high velocity), whereas in the asthenosphere, the absorption peak occurs at a frequency below that of seismic shear waves. The absorption band behavior has been documented in laboratory studies (Jackson and Faul, 2010; Sundberg and Cooper, 2010). Further study will be needed to confirm the suggested behavior of the absorption peak.

As described earlier, radial and azimuthal seismic anisotropies of the uppermost mantle are well documented. Azimuthal anisotropy may change direction with depth: fast direction in the lithosphere and asthenosphere in the direction of plate spreading and the direction of plate motion, respectively. It remains to be understood whether changes in the strength or direction of anisotropy with depth can be sufficiently rapid to explain the G discontinuity (e.g., Fischer et al., 2010).

7.08.2.5 Upper Mantle Electrical Conductivity: Water Content and Temperature

Long-period magnetotelluric studies of mantle electrical structure beneath the eastern North Pacific Ocean can detect electrical conductivity structure between 150 and 1000 km (Lizarralde et al., 1995). Interpretation of these data reveals a conductive zone between 150 and 400 km in depth with continuously decreasing conductivity at greater depth. Mantle conductivity in this region is comparable to that in the Basin and Range Province and much higher than that in the Canadian Shield. High conductivities could be explained by the presence of gravitationally stable partial melt. Alternatively, conductivity estimates on measurements of hydrogen solubility and diffusivity in olivine can explain the high conductivities observed. This reinforces the possible role of water in controlling the physical properties, and particularly the rheology, of the upper mantle (Hirth and Kohlstedt, 1996; Hirth et al., 2000). Several recent studies measuring the electrical conductivity of wet olivine (Wang et al., 2006; Yoshino et al., 2006) reach opposing conclusions on whether intragranular water (H defects) alone can explain observed conductivities or whether partial melt is required.

7.08.2.6 The Age of Thermal Convective Instability Beneath Oceanic Lithosphere

Global gravity anomaly maps (Jekeli, Vol. 3) derived from Seasat altimetry data display gravity lineations in the Pacific with 150–200 km wavelength aligned in the direction of plate motion (Haxby and Weissel, 1986) as shown in **Figure 5**. Haxby and Weissel (1986) suggested that these gravity lineations were a consequence of mantle convection currents organized by plate motions. In a sheared fluid layer with an initial linear temperature gradient, corresponding to heating from below and cooling from above, convective instability takes the form of convective rolls aligned with plate motion (Richter, 1973; Richter and Parsons, 1975). The presence of gravity lineations due to convective instability beneath lithosphere only a few million years old contrasted with the view that convective instability at ages ~70 My explained the flattening of old seafloor.

Motivated by the observed gravity anomalies, numerical studies of steady-state finite amplitude thermal convection showed that convective instability at young ages could be consistent with mantle thermally activated creep rheology (Buck, 1985) and estimated the convective heat fluxes required to explain the observed gravity anomalies (Lin and Parmentier, 1985). Numerical solutions for the development of thermal convection in a fluid layer cooled from above, with a plausible range of rheological parameters, predicted the effect of convective instability on geophysical observables (Buck and Parmentier, 1986). This study assumed that stresses generated by cool mantle sinking from the bottom of the unstable thermal boundary layer did not contribute to seafloor topography and resulting geoid anomalies, implying that this negatively buoyant mantle was supported by higher-viscosity mantle at depth. Robinson and Parsons (1988) showed that this is a reasonable approximation if mantle viscosity increases sufficiently with depth. If this were not so, convective instability, which increases the rate of cooling of mantle columns, would cause more rapid seafloor subsidence rather than reduced subsidence as previously assumed (cf. O'Connell and Hager, 1980). The calculated geoid anomaly showed a rapid decline in the geoid–age slope as convective instability developed and compared favorably to observed variations in geoid slope with age as shown in **Figure 3**. Other more recent studies argue that convective motions that develop at young ages may explain depth and geoid–age data better than plate model advocates have assumed. Doin and Fleitout (1996) showed that a uniform heat flux supplied to the bottom of the thermal boundary layer, presumably by convective motions, explains depth–age

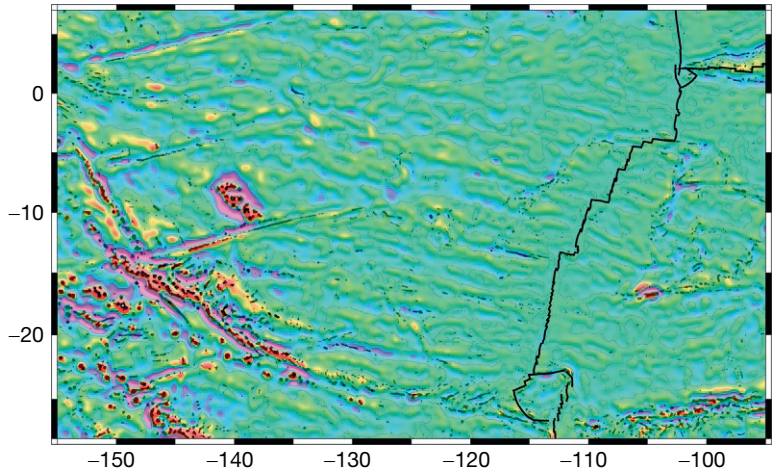


Figure 5 Haxby gravity lineations depicted in this satellite gravity image by Sandwell and Fialko (2004). Band-pass filtered gravity anomaly ($80 < \lambda < 600$ km) derived from retracked satellite altimeter data. Color scale saturates at ± 15 mGal. Gravity lineations with 140 km wavelength develop between the ridge axis and 6 Ma and are oriented in the direction of absolute plate motion. Lineations on older seafloor have somewhat longer wavelength (~180 km) and cross the grain of the seafloor spreading fabric. Gravity lineations also occur on the Nazca Plate to the east of the East Pacific Rise.

and plate model and also the rapid decline in the geoid–age slope at young ages.

Haxby and Weissel (1986) suggested the presence of longer wavelengths of small-scale convection beneath older regions of the Pacific. Wessel et al. (1996) identified gravity undulations with wavelengths of ~280 and 1000 km that form small circles about the current pole of Pacific Plate motion. The shorter wavelengths would correspond to the Haxby gravity lineations. The longer wavelengths could reflect the spacing of hot spot tracks or a larger scale of thermal boundary convective instability corresponding to that usually invoked to explain the plate model. The earlier study of Cazenave et al. (1995) had also indicated a wavelength of about 1000 km. Mantle seismic tomography using ray paths along a profile connecting Tonga and Hawaii identified seismic velocity anomalies with a spacing of about 1500 km that correlated with variations in gravity and bathymetry (Katzman et al., 1998). As the depth of a convectively cooled mantle layer increases with time, simple dimensional and scaling arguments suggest that the wavelength of convective motions should increase, a behavior seen in early studies of convective instability (Buck and Parmentier, 1986), but more recent studies, summarized in the succeeding text, further address this.

The Haxby gravity lineations are subtle features of the gravity field, generally much smaller than gravity anomalies due to fracture zones, for example. The current Pacific and Nazca Plate motion is oblique to fracture zones so that lineations aligned with plate motion cut obliquely across them. For other plates that move nearly parallel to fracture zones within them, gravity lineations of comparable amplitude to those in the Pacific would be difficult to detect, so that to identify convective instability at young ages requires other evidence. If convective instability occurs beneath the fast moving Nazca and Pacific Plates, does it also occur beneath more slowly moving plates created by slower spreading? As discussed earlier, Sandwell and Schubert (1980) found that geoid height for the Atlantic and southeast Indian spreading centers decreases approximately linearly with age with a geoid–age slope comparable to that expected for purely conductive cooling. This suggests that small-scale convective instability at young ages does not occur beneath these more slowly spreading plates.

Previous studies have often treated either hot spots or small-scale convective instability as the mechanism of lithospheric heating in all ocean basins. However, one mechanism of heat transfer need not be responsible for heating the lithosphere everywhere. With due caution concerning the nonlinearity of thermal convection (meaning that simply adding heat fluxes due to hot spots and small-scale convection may not be valid), it may even be reasonable to think that both mechanisms operate simultaneously but to different degrees beneath different ocean basins. Gravity lineations at young ages are visible only on the Pacific and Nazca Plates, suggesting that convective instability at young ages may develop under these fast moving plates. In the South Atlantic, where hot spot tracks are sparse and gravity lineations are not detectable, square root of age subsidence to ages exceeding 100 My would be consistent with purely conductive mantle cooling beneath this slower moving plate.

7.08.2.7 Implications from Recent Theoretical and Experimental Studies of Convective Instability

Convective instability of a thermal boundary layer has been examined in numerous studies using a range of analytic methods (see, e.g., Yuen and Fleitout, 1984; Marquart et al., 1999) and both laboratory (00003, this volume) and numerical experiments (00003 and 00005, this volume; Zaranek and Parmentier, 2004), the latter taking advantage of advances in computer speed and memory, to better understand the development of convective instability in a time-dependent basic state (conductive cooling) and temperature-dependent viscosity. The strong temperature dependence of mantle viscosity presents a significant challenge for numerical experiments and emphasizes the important continuing role of laboratory experiments.

Davaille and Jaupart (1994) derived scaling laws for convective onset time (age) and heat flow in a viscous fluid with strongly temperature-dependent viscosity cooled from above.

A convective heat flux f independent of the depth of the convecting fluid, implying that the boundary thickness is small compared to the fluid depth, is given by

$$f = Ck\left(\frac{\rho\alpha g}{\mu\kappa}\right)^{1/3}\Delta T_c^{4/3}$$

where k, κ, α, and ρ are the thermal conductivity, thermal diffusivity, thermal expansion coefficient, and density, respectively, and $C = 0.16$ is a dimensionless constant determined from laboratory or numerical experiments. This expression for the heat flux follows directly from dimensional analysis. Davaille and Jaupart (1994) showed that the convecting thermal boundary layer coincided with a tenfold increase in viscosity, corresponding to temperature decrease ΔT_c. For thermally activated creep,

$$\mu = \mu_m \exp\left[\frac{Q}{R}\left(\frac{1}{T} - \frac{1}{T_m}\right)\right]$$

where μ_m and T_m correspond to the temperature and viscosity beneath the thermal boundary layer. Then,

$$\Delta T_c = 2.24\frac{RT_m^2}{Q}$$

This heat flux scaling and the values of C are also confirmed by numerical experiments (Grasset and Parmentier, 1998). Heat flux as a function of μ_m and Q is shown in Figure 6. In the mantle, a heat flux due to small-scale convection comparable to that which would be required by the plate model ($\sim$40 mW m^{-2}) and creep activation energies from $Q = 250$ to 600 kJ mol^{-1} would require asthenosphere viscosities between 3×10^{18} and 4×10^{17} Pa s.

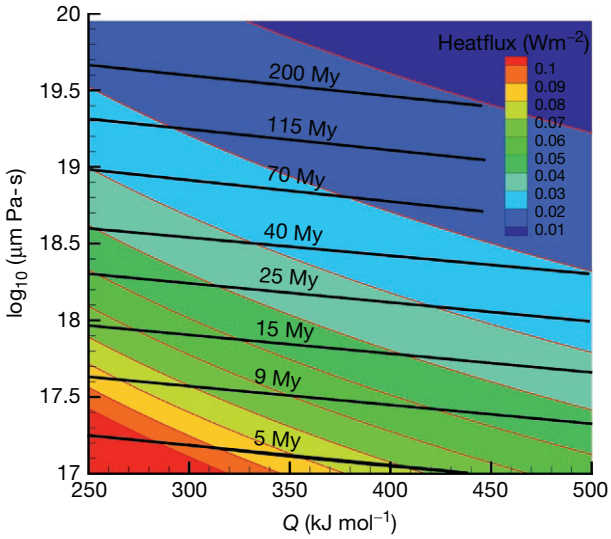


Figure 6 Contours of mantle heat flux beneath old oceanic lithosphere (based on heat flux scaling of Davaille and Jaupart (1994)) and onset time (age) of convective instability (reproduced from Zaranek SE and Parmentier EM (2004) Convective instability of a fluid with temperature-dependent viscosity cooled from above. *Earth and Planetary Science Letters* 224: 371–386) as a function of mantle viscosity and activation energy for thermally activated creep. For mantle heat flow below old oceanic crust of $\sim$50 mW m^{-2} (see Figure 2), onset of convection at young ages should occur if the creep activation energy is sufficiently high and mantle viscosity sufficiently low.

The small scale of thermal structures and convective motions in the mantle predicted by laboratory and numerical experiments is important to appreciate. For a convective heat flux beneath old lithosphere of $\sim$40 mW m^{-2} and $\Delta T_c \sim 100\,^{\circ}$C, the convecting thermal boundary layer thickness $\delta_c \sim k\Delta T_c/f$ is < 10 km, and convective instability of the boundary layer is expected to generate cold downwellings of comparable width.

The onset time of convective instability can also be determined from both laboratory and numerical experiments. For $Q = 250$ kJ mol^{-1} and a mantle temperature of 1300 °C, Davaille and Jaupart predicted the onset time from their laboratory-derived scaling to be in the range 52–65 My. High-resolution numerical experiments examining the onset of convection in a fluid with strongly temperature-dependent viscosity have been reported in several recent studies (Huang et al., 2003; Korenaga and Jordan, 2003a; Zaranek and Parmentier, 2004). These studies predict shorter onset times than earlier studies. Since convective motions, even those restricted to the upper mantle, must allow for scales as large as 1000 km, the small scales of convective instability indicated earlier would be difficult to resolve numerically. An understanding of the scaling of onset time with temperature dependence is needed to extrapolate well-resolved numerical experiments to the very stronger temperature dependence of viscosity expected in the mantle. Recent numerical experiments, which are all in good agreement with each other, indicate a scaling for onset time that differed from that proposed by Davaille and Jaupart (2004). The onset time scaling of Zaranek and Parmentier (2004), for example, leads to the predictions shown in Figure 6.

In Figure 6, contours of heat flux provided to the base of the conductive lithosphere by small-scale convection in color shading and onset time in solid contours are plotted as a function of the creep activation energy Q and mantle viscosity μ_m beneath the thermal boundary layer. The upper mantle viscosity beneath spreading centers and areas of active plate extension such as the North American Basin and Range Province is not expected to exceed about 10^{19} Pa s (Bills, 1994; Passey, 1981; Sigmundsson, 1991). Estimates based on laboratory rheological measurements suggest an even lower viscosity of 10^{18} Pa s (Hirth and Kohlstedt, 1996). If small-scale convection provides all the heat flux to old seafloor ($\sim$45–50 mW m^{-2}; see Figure 2) and if an upper mantle viscosity of 10^{18} Pa s is assumed, convective instability would occur at ages of $\sim$15 My for $Q = 350$–400 kJ mol^{-1}, values intermediate between diffusion ($\sim$300 kJ mol^{-1}) and dislocation creep ($\sim$500 kJ mol^{-1}) in olivine.

The onset of convection at ages $\leq$10 My would require μ_m below 10^{18} Pa s and Q near the value for dislocation creep in olivine. For convective instability at 70 My with a convective heat flux of 40 mW m^{-2}, a value of μ_m higher than those cited earlier would be required with a creep activation energy lower than 300 kJ mol^{-1}. For Q within the 300–500 kJ mol^{-1} range, onset of convection in the 50–70 My age range would be

possible only with an upper mantle viscosity of 10^{19} Pa s and then only if the convective heat flux due to small-scale convection beneath old seafloor were $\leq$30 mW m^{-2}. This may be possible if small-scale convection transports only a fraction of the heat needed to explain the surface heat flow and bathymetry of old seafloor, the remainder being due to heating of the lithosphere at hot spots, as discussed earlier.

Mantle viscosity μ_m may differ near a spreading center and beneath older seafloor. Near spreading centers, the presence of melt may reduce viscosity. Conversely, since the presence of intragranular water in nominally anhydrous minerals reduces creep viscosity and water is a relatively incompatible element, the extraction of melt may dehydrate the mantle, thus increasing viscosity (Hirth and Kohlstedt, 1996; Karato, 1986). As discussed in the succeeding text, the latter effect may be the more dominant one, depending on the sensitivity of creep rates to water content and the water content of the mantle prior to melting.

As indicated earlier, convective instability at young ages may be present only beneath the Nazca and Pacific Plates, which are both fast moving and formed at a high spreading rate. Why might convective instability at young ages be restricted to fast spreading or fast moving plates? Is spreading rate or plate velocity the determining factor? Slower spreading rates should lead to a thicker thermal boundary layer beneath the spreading axis, thus promoting convective instability at younger ages (Sparks and Parmentier, 1993). Other possible effects of spreading rate are discussed later. The linear stability analysis of Korenaga and Jordan (2003b) indicates that convective rolls aligned with plate motion may be stable only in the presence of strong shearing beneath fast moving plates. However, rheological effects may also be important. Beneath fast moving plates, higher strain rates could create smaller grain size resulting in more rapid diffusion creep with lower μ_m or higher stresses resulting in deformation dominated by dislocation creep with higher Q. As shown in **Figure 6**, both lower μ_m and higher Q would favor convective instability at younger ages. Rising mantle that is enriched in water relative to that in other regions might also explain a lower mantle viscosity beneath the EPR, thus promoting early onset times.

The scale or wavelength of small-scale convection appears to increase to $\geq$1000 km beneath the old Pacific seafloor. Korenaga and Jordan (2004) recently examined how the horizontal scale of convective motions in a fluid with temperature- and depth-dependent viscosity would increase with age, including in particular the effect of an endothermic phase transition at the base of the upper mantle. They suggested that convection initiated at small scales beneath moving plates could eventually penetrate the stable phase transition and evolve into whole mantle convection.

All of the studies discussed earlier treat the development and evolution of convective motions as two-dimensional (2D) and time-dependent, most closely approximating convective motions in a vertical plane orthogonal to plate motion as a function of plate age. Despite significant advances in both computer speed and numerical methodologies, fully three-dimensional treatments of thermal convection with the very strongly temperature-dependent viscosities believed to characterize mantle flow remain challenging (see 00005, this volume). Numerical solutions must simultaneously resolve 10 km thick thermal boundary layer layers and cold plumes embedded in several thousand kilometer-scale flow.

van Hunen et al. (2003) studied 3D convective instability beneath a moving plate. In agreement with earlier analysis, plate motion enhanced the development of longitudinal convective rolls relative to motion-perpendicular transverse rolls. They found that the onset age of fully 3D convective motions was similar to that in earlier 2D studies. It would be interesting to more closely examine the evolution of convective motions with age, in particular how the scale or wavelength of convective motions increases as cooling proceeds. A first impression based on van Hunen et al. (2003) (**Figure 1**) suggests that the scale of convective motions does not increase as rapidly as suggested by earlier 2D studies.

7.08.3 Convective Instability and Melting in the Upper Mantle

7.08.3.1 Intraplate Volcanism as an Indicator of Upper Mantle Convective Activity

Volcanism away from plate boundaries (00009, this volume; McNutt, Vol. 1) is generally thought to be a consequence of decompression melting due to convective motions that arise from deeper in the mantle or from instability generated in the upper mantle. Linear, long-lived, and age-progressive volcanic chains have been explained as the manifestation of fixed hot spots, possibly generated by buoyant plumes of rising material (00009, this volume) originating deep in the mantle (Morgan, 1971). While the Wilson–Morgan fixed hot spot model has been successfully applied to volcanic chains in all ocean basins, important disagreements have been recognized between geochronological observations and the simplest predictions of the model. For example, Okal and Batiza (1987) and McNutt (1998) outlined numerous examples where the fixed hot spot model fails to explain (a) observed departures from linearity of individual volcanic chains and inconsistent orientations among multiple chains that lie on the same plate, (b) short-lived chains and ones that fluctuate in size, and (c) violations of predicted age-progressive behavior. For example, the Cook–Austral Chain (Turner and Jarrard, 1982) and the Line Islands (Schlanger et al., 1984) exhibit complex age progressions with volcanic activity occurring along multiple lineaments and with volcanoes of distinctly different ages in close proximity to one another. Furthermore, some linear volcanic ridges in the Pacific form much more rapidly than would be predicted by fixed hot spot model (Bonatti et al., 1977; Sandwell et al., 1995).

The creation of large igneous provinces (00010, this volume) or oceanic plateaux by partial melting of the starting plume head is a corollary of the mantle plume hypothesis for the origin of hot spots (Richards et al., 1989). Of the 14 possible Pacific hot spot tracks studied by Clouard and Bonneville (2001), only 3 (Louisville, Easter, and Marquesas) can be traced to an oceanic plateau. Seven hot spots have short tracks $<$35 My and clearly cannot be traced to an oceanic plateau. Koppers et al. (2003) found that linear volcanic chains in the western Pacific and South Pacific Superswell region typically display intermittent volcanic activity with longevities shorter than 40 My, superposed volcanism, and motion

relative to other longer-lived hot spots. Finally, study of marine satellite gravity data in the Pacific (Wessel and Lyons, 1997) shows the presence of many volcanoes with a range of sizes in a variety of geologic settings. Most of these volcanoes do not clearly align in chains or ridges, and most seem too small to be explained by deep mantle plumes. While the long-lived age-progressive volcanic chains that fit the fixed hot spot model are remarkable features, other mechanisms of intraplate volcanism must also be active (see also 00010, this volume).

7.08.3.2 Other Possible Mechanisms for Intraplate Volcanism

Diffuse plate extension (Sandwell et al., 1995) has been frequently cited as a possible mechanism of intraplate volcanism. Proposed partly on morphological grounds, this mechanism was thought to explain rapid propagation of volcanic ridges and the formation of volcanic ridges in troughs of topography associated with the Haxby gravity lineations. Volcanism was thought to be attributable either to allowing melt already present in the upper mantle access to the surface or to decompression melting in upwellings associated with lithospheric boudinage. Dunbar and Sandwell (1988) calculated that 10% extension would be required for this mechanism. However, the amount of extension that can be allowed in the Pacific is $<1\%$ (Gans et al., 2003; Goodwillie and Parsons, 1992), seemingly much too small for significant boudinage. Thus, agreement seems to be emerging that the boudinage hypothesis does not satisfy available observations. An interesting alternative is cracking of the lithosphere under the action of thermal contraction bending moments (Gans et al., 2003; Sandwell and Fialko, 2004). In this hypothesis, cracking of the plate allows lithosphere between cracks to bend, thus relieving thermal bending moments. Cracks form topographic troughs, and magma already existing beneath the lithosphere exploits these regions to reach the seafloor. Mantle seismic velocities beneath volcanic ridges formed in this way should be higher than in adjacent mantle, which does not seem to be the case, at least beneath the volcanic ridges where regional seismic data are available. Weeraratne et al. (2007) found lower seismic velocities, implying higher temperatures and/or more melt beneath these volcanic ridges. However, the ridges examined in this study are all near the East Pacific Rise spreading center on seafloor younger than 3 Ma. Seismic velocities beneath recently active volcanic ridges further from the spreading axis have not yet been studied in comparable ways.

7.08.3.2.1 Melting in convectively driven upwellings

Volcanism due to decompression melting in small-scale convective upwellings is perhaps the favored popular hypothesis for the origin of volcanic ridges aligned with plate motion (see also 00010, this volume). Melting in convective upwellings would be consistent with lower seismic velocities in the mantle beneath ridges. Buoyant convective upwellings should elevate the seafloor so that the resulting volcanism should occur on topographic highs. However, volcanic ridges were observed to lie in lows of the gravity lineations. If gravity lows correspond to topographic troughs, this would be inconsistent with a convective origin (Sandwell et al., 1995). However, seafloor topography with 100 m amplitudes over horizontal scales of several hundred kilometers is a subtle feature compared with the topography of volcanic constructs and the flexure caused by this loading. After removing lithospheric flexure due to the topographic loading, Harmon et al. (2006) found that volcanic ridges near the EPR actually lie on topographic highs. Topographic highs correspond to negative residual mantle Bouguer gravity anomalies that were obtained from the observed free-air anomaly by subtracting the effect of crustal thickness variations and the attraction of topography. If this relationship holds for other volcanic ridges, particularly those further from the EPR axis, then a convective origin for volcanic ridges and gravity lineations would be indicated. Harmon et al. (2011) argued for a convective origin based on the combined analysis of seismic velocity and density variations in the mantle.

7.08.3.2.2 Melting due to shear-driven upwelling

Decompression melting has frequently been associated with convectively driven mantle upwelling as discussed earlier. However, mantle upwelling due to plate motions and asthenospheric shear may also be driven by variation in plate thickness or by viscosity variations in a sheared asthenosphere. The fracture zones, across which seafloor age and thermal boundary layer thickness change abruptly, often move obliquely over the underlying mantle, providing one example of a setting in which upwelling can result from changes in plate thickness. Till et al. (2010) explored a recent model of this process in the context of a continental margin. Conrad et al. (2010) also presented models of mantle upwelling that results from lithosphere thickness variations.

Conrad et al (2010) also suggested that mantle upwelling may also result from viscosity variations in a sheared asthenosphere. Viscosity variations in the asthenosphere may be a consequence of preexisting variations in mantle composition, for example, water content, or the presence of melt. Conrad et al. (2011) presented evidence for a correlation between intraplate volcanic activity and rapid shearing of the asthenosphere. It is interesting to wonder how such viscosity variations in the asthenosphere might arise and the timescale on which they can be preserved. Volcanic chains near the East Pacific Rise spreading center, often attributed to upwelling in convective instability, provide an interesting example in which the space–time variation of volcanism may be due to the presence of fingers of low-viscosity mantle (Weeraratne et al., 2007) and upwelling due to asthenospheric shear in the presence of these viscosity variations (Ballmer et al., 2013). It is interesting to wonder if viscosity variations might arise spontaneously due to feedback between viscosity and melting.

7.08.3.3 Buoyant Decompression Melting

In addition to decompression melting in thermally driven convection, intraplate volcanism may be a product of decompression melting in convective upwellings that result from the buoyancy associated with melting itself (Hernlund and Tackley, 2003; Raddick et al., 2002; Tackley and Stevenson, 1993). This 'buoyant decompression melting' in a layer that is initially at its melting temperature may self-organize from small, initially random perturbations and thus might be termed 'magmatic convective storms.' Melt extraction leaves

behind a buoyant and creep-resistant residual mantle. The accumulation of this relatively immobile and infertile mantle ultimately limits the amount of melt that can be produced by the decompression melting mechanism (Raddick et al., 2002). This results in an inverse correlation between the rate of melt production and the duration of melting, which may be diagnostic of the buoyant decompression melting process.

Buoyant decompression melting could occur in a number of geologic settings; however, if the oceanic upper mantle has previously melted beneath a spreading center and subsequently cooled, spontaneous buoyant decompression melting may be possible only in regions where a large-scale mantle upwelling can counteract conductive cooling, keeping the mantle at its solidus temperature over some depth range. The South Pacific Superswell has been interpreted as a region of large-scale mantle upwelling (Gaboret et al., 2003; McNutt, 1998), and buoyant decompression melting may explain the abundance of volcanism associated with it. Alternatively, to balance the downward flux associated with sinking lithosphere, a small upward velocity should be present in large areas of the upper mantle away from convergent plate boundaries. Since transition zone mineral phases like the β- and γ-olivine have a larger water storage capacity than their lower pressure isomorph, upwelling mantle may dehydrate forming a water-rich melt above the transition zone (e.g., Bercovici and Karato, 2003). During the upward percolation of buoyant melt, as envisioned in an early study by Frank (1968), buoyant instabilities may lead to localization of melting, upwelling, and surface volcanism.

In the absence of a large-scale upwelling, where mantle is slightly cooler than its melting temperature, buoyant melting may not occur spontaneously but may be triggered by some initial upwelling due to relief on the bottom of the lithosphere, for example, across oceanic fracture zones that move obliquely across the mantle (e.g., Raddick et al, 2002). This may provide a physical explanation for intraplate volcanism that is controlled by lithosphere structure.

7.08.4 Upwelling and Melting Beneath Oceanic Spreading Centers

7.08.4.1 Melting, Melt Extraction, and the Chemical Lithosphere

Mantle upwelling and decompression melting beneath spreading centers (Jian Lin, Vol. 6) are expected to have important consequences for the development of convective motions as the lithosphere ages. Two effects may be particularly important: (a) melting and melt extraction are expected to affect both rheology and density and (b) convective instability beneath the spreading axis due to density variations associated with the presence of melt and its extraction from residual mantle may influence the scale and evolution of off-axis convective instability. Forsyth (1992) provided a relatively recent review of mantle flow beneath spreading centers; and the MELT Experiment (Forsyth et al., 1998) provides the best geophysical evidence on the amount and distribution of melt in the mantle beneath a very fast spreading section of the EPR. Geochemistry, for example, Lu–Hf systematics and U–Th disequilibrium, provides evidence that garnet was present during melting, thus providing a minimum estimate of ~70 km for the beginning of melting in upwelling mantle.

Hirth and Kohlstedt (1996) and Phipps Morgan (1997) pointed out that melting beneath spreading centers should produce a compositional lithosphere that is both more viscous, as mentioned earlier, and compositionally buoyant. Intragranular water that enhances the creep rate of nominally anhydrous mantle minerals behaves as a highly incompatible element during melting. The extraction of melt during fractional melting would leave behind a dry residual mantle. Removing essentially all the water present at concentrations inferred for upper mantle that melts beneath spreading centers may increase its viscosity by more than a factor of 100 (Hirth and Kohlstedt, 1996) (Kohlstedt, Vol. 2). The thickness of the residual layer should be comparable to the maximum depth of melting beneath a spreading center. Evans et al. (2005), using electrical conductivity, inferred a 60 km dehydration depth beneath a fast spreading section of the EPR (the MELT area at 17° S). The dehydrated residual mantle layer generated at a spreading center might be thought to correspond to the plate in the plate model seafloor evolution. However, this depth of dehydration appears significantly thinner than estimates of plate thickness in the range of 95–125 km discussed earlier. One possibility is that small amounts of wet melt form at the even greater depths comparable to the plate thickness. Interpretations of seismic data from the MELT Experiment do suggest small amounts of melting at depths exceeding 100 km (Forsyth et al., 1998).

7.08.4.2 Influence of Melt Extraction Mechanism

The rheological consequences of melt extraction should depend on the mechanism of melt migration. Does melt percolating upward remain distributed along mineral grain edges maintaining equilibrium with solid mantle as it goes? Or does it localize into larger channels after only small amounts of melt form? Melt that rapidly localizes into larger channels approximates ideal fractional melting that removes incompatible elements more effectively than equilibrium transport.

A fundamental observation is that the chemical composition of basalt erupted at spreading centers is not in equilibrium with residual mantle at low pressure (e.g., Elthon and Scarfe, 1984; O'Hara, 1965; Stolper, 1980). Melts must rise from depths of at least 30 km to the surface without extensive re-equilibration with surrounding mantle at shallow depth in order to preserve this deep geochemical signature. Focused melt flow through high-permeability conduits or channels is one possible mechanism. Several lines of evidence suggest that the porosity structure in the melt generation and extraction region of the mantle is heterogeneous consisting of interconnected high-porosity dunite channels embedded in a low-porosity harzburgite or lherzolite matrix (Kelemen et al., 1997 and references therein). Remnant dunite channels have been observed as veins, tabular or sometimes irregular-shaped bodies in ophiolites. These dunite dikes or veins make up 5–15% of the mantle in the Oman ophiolite with widths ranging from tens of millimeters to ~200 m and length from tens of meters to at least 10 km (Kelemen et al., 1997). Although their spatial distribution in the mantle is still not well constrained, the presence of a high-porosity,

interconnected, coalescing network of dunite channels present above or within the melting region has been envisioned. Porous flow through dunite channels is capable of producing observed Uranium-series disequilibriums and significant trace element fractionations during mantle melting and melt extraction (e.g., Kelemen et al., 1997; Lundstrom, 2003; Spiegelman and Elliott, 1993; Spiegelman and Kelemen, 2003).

At fast spreading rates, mantle upwelling and melting appear to occur over a several hundred kilometer wide region (e.g., Forsyth et al., 1998), but the oceanic crust is emplaced within a few kilometers of the spreading axis. Thus, a second major constraint on the mechanism of melt migration is that it must be capable of focusing melt to the spreading axis. If the viscosity of upwelling mantle is sufficiently high ($\geq 10^{20}$ Pa s), pressure gradients in the mantle flow may be sufficient to drive melt to the spreading axis (Phipps Morgan, 1987; Spiegelman and McKenzie, 1987). Alternatively, melt may migrate vertically to collect in a decompaction boundary layer that develops as melt begins to freeze. This should create a high-porosity melt channel that slopes away from the spreading axis providing a conduit for melt flow toward the axis (Sparks and Parmentier, 1991; Spiegelman, 1993). The first full numerical simulation of mantle upwelling melting and melt migration beneath a spreading center showing this decompaction boundary layer is shown in **Figure 7(a)** from Katz (2010). Rabinowicz and Ceuleneer (2005) suggested, in fact, that dunites in the Oman ophiolite may be the preserved remnants of this decompaction boundary layer. If melt in the decompaction layer is isolated in dunite channels, then some signature of high pressures will be preserved in the melts that accumulate near the spreading axis. However, melting need not be perfectly fractional so that mantle need not be fully dehydrated.

No seismic or electromagnetic evidence on the distribution of melting and upwelling in the mantle comparable to that for southern EPR is available for slower spreading centers. If buoyancy related to melting is important, then 'active' upwelling could lead to more localized melting at slow spreading rates. Active upwelling would produce higher, more uniform degrees of melting creating more strongly and uniformly dehydrated residual mantle less prone to convective instability. This is illustrated in **Figure 7** from Braun et al. (2000).

7.08.4.3 2D Versus 3D Upwelling and the Spreading Rate Dependence of Seafloor Structure

At slow spreading rates, buoyancy related to melting may also result in localized columns of upwelling and melting beneath the spreading axis (Choblet and Parmentier, 2001; Parmentier and Phipps Morgan, 1990). At high spreading rates, upwelling remains sheetlike along the spreading axis. This is one possible explanation for the difference in seafloor morphology and structure between fast and slow spreading centers and may also explain the strongly lineated morphology of seafloor produced at slow spreading rates (Phipps Morgan and Parmentier, 1995). At high spreading rates, buoyant decompression melting may produce off-axis upwelling columns that may explain near-axis volcanic ridges (Jha et al., 1997), providing a possible alternative to melting in thermally driven upwellings as discussed earlier.

7.08.5 Summary

The deviation of the age dependence of seafloor depth from predictions of a simple conductive thermal boundary layer (half-space cooling model) need not be explained by only a single mechanism. Both heating by hot spots and convective instability due to cooling from above may affect thermal evolution but differ in importance in different settings, for example, the North Atlantic and Pacific, respectively. Discussion has sometimes focused on which mechanism is the correct one rather than on assessing which may be more important and why. The relative importance of convective instability should depend on spreading rate and mantle composition, perhaps particularly through the effect of water on rheology.

The seafloor age at which convective instability begins clearly depends strongly on spreading rate and particularly mantle rheology. Convective instability at young ages remains a preferred explanation for gravity lineations on the Pacific and Nazca Plates but is not required beneath other plates. Beneath thicker, colder lithosphere at slow spreading centers, small-scale convection may be present but less visible. If small-scale convection is present beneath the fast spreading EPR but not beneath slower spreading centers, a number of explanations for this possible difference in behavior can be envisioned. The higher shear rates beneath rapid moving plates may organize convection into a lineated structure that is more visible in gravity data than other less organized patterns. Higher shear rates may also result in lower effective viscosity of the upper mantle beneath faster moving plates, therefore yielding earlier onset times.

The reduced rate of old seafloor has frequently been attributed to convective instability of the thermal boundary layer beneath a thickening conductive lid. If so, then convective instability would occur at ages of $\geq$70 My, which maintains a nearly uniform mantle temperature beneath the conductive lid. If elastic wave speed is primarily a function of temperature, an upper mantle temperature that does not decrease with seafloor age does not explain an elastic shear wave velocity, which seismic tomography indicates continues to increase with age. If elastic shear wave speed is reduced by the presence of intracrystalline water in nominally anhydrous mantle minerals, then small degrees of partial melting could increase seismic velocity through the extraction of water, which behaves as a relatively incompatible trace element. The presence of melt has also been suggested to explain the G discontinuity, which some studies suggest to be widespread in the oceanic upper mantle. However, other mechanisms to create the G discontinuity are under active consideration. The continuing increase in seismic velocity beneath seafloor with an old age and the presence of the G discontinuity seem likely to have important rheological implications that could improve our understanding of oceanic upper mantle convective instability.

Intraplate volcanism not associated with hot spots on the Pacific and Nazca Plates is much more abundant than in other ocean basins. Does this reflect an influence of deeper mantle processes or convective instability that develops within the upper mantle? The latter possibility would be broadly consistent with a lower-viscosity (higher strain rate, smaller grain size, and/or wetter) mantle beneath the Pacific and Nazca Plates than elsewhere in the upper mantle.

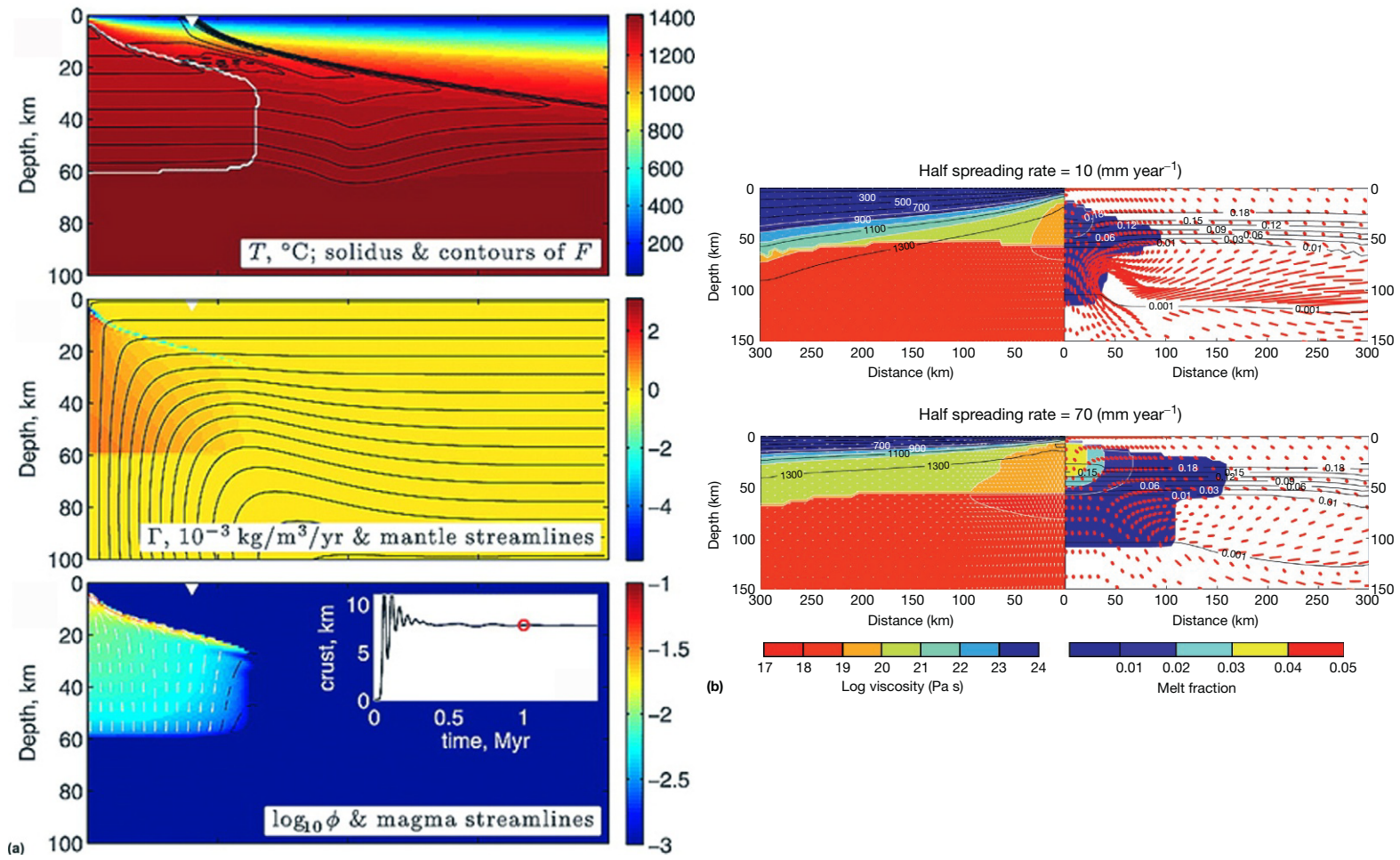


Figure 7 (a) Numerical simulation of mantle flow, melting, and melt migration beneath a spreading center with half spreading rate 4 cm year^{-1}, a mantle viscosity of 10^{18} Pa s, and potential temperature of 1375 °C from Katz (2010) illustrating the presence of a decompaction boundary layer. (top) Temperature, (middle) rate of melting, and (bottom) melt fraction present obtained from the simulation after 1 Ma of model time. White triangles show the distance of plate motion over this period. Black curves in top panel connect points of constant degree of melting, starting with 1% at the bottom of the melting region and increasing in steps of 2.5%; these lines cluster along the edge of the region that has experienced melting since spreading initiated. In this model, the maximum degree of melting is 20%. The inset in the bottom panel shows crustal thickness as a function of time. The red point marks the crustal thickness at 1 Ma. (b) Mantle upwelling and melting beneath fast and slow spreading centers from Braun et al. (2000). Solid contours in the left panel show the mantle viscosity, and black lines depict isotherms (at 200 °C intervals) for a potential temperature of 1350° and an adiabatic gradient of 0.3°C km^{-1}. Ellipses (red ovals) representing accumulated finite strain are plotted along solid flow streamlines corresponding to the velocity field shown by white vectors in the left panel. Solid contours on the right panel show the steady-state melt fraction (>0.1%) overlain by isodepletion contours (black lines). Rheology includes estimates of the effects of dehydration, melt, and grain boundary sliding with a reference viscosity of 5×10^{18} Pa s. Since the effects of dehydration dominate those of melt, the viscosity and velocity structures resemble more passive-like plate spreading flow above the dry solidus. Even with lower reference viscosities, buoyant flow is restricted to depths beneath the dry solidus. Buoyant localization of solid flow and melt generation at depth increases as the half spreading rate decreases. The magnitude and localization of strain also increase with decreasing spreading rate.

References

Ballmer MD, Conrad CP, Smith EI, and Harmon N (2013) Non-hotspot volcano chains produced by migration of shear-driven upwelling toward the East Pacific Rise. *Geology* 41: 479–482. http://dx.doi.org/10.1130/G33804.1.

Bercovici D and Karato S (2003) Whole-mantle convection and the transition-zone water filter. *Nature* 425: 39–44. http://dx.doi.org/10.1038/nature01918.

Bills BG (1994) Viscosity estimates for the crust and upper mantle from patterns of lacustrine shoreline deformation in the eastern Great Basin. *Journal of Geophysical Research* 99: 22059–22086.

Bonatti E, Harrison CGA, Fisher DE, et al. (1977) Easter volcanic chain (southeast Pacific): A mantle hotline. *Journal of Geophysical Research* 82: 2437–2478.

Braun M, Hirth G, and Parmentier EM (2000) Effects of deep damp melting on mantle flow and melt generation beneath mid-ocean ridges. *Earth and Planetary Science Letters* 176: 339–356.

Buck WR (1985) When does small scale convection begin beneath oceanic lithosphere? *Nature* 313: 775–777.

Buck WR (1987) Analysis of the cooling of a variable-viscosity fluid with applications to the Earth. *Geophysical Journal of the Royal Astronomical Society* 89: 549–577.

Buck WR and Parmentier EM (1986) Convection beneath young oceanic lithosphere: Implications for thermal structure and gravity. *Journal of Geophysical Research* 91: 1961–1974.

Carlson RL and Johnson HP (1994) On modeling the thermal evolution of the oceanic upper mantle: An assessment of the cooling plate model. *Journal of Geophysical Research* 99(B2): 3201–3214. http://dx.doi.org/10.1029/93JB02696.

Cazenave A (1984) Thermal cooling of the oceanic lithosphere: New constraints from geoid height data. *Earth and Planetary Science Letters* 70: 395–406.

Cazenave A, Lago B, and Dominh K (1982) Geoid anomalies over the north-east Pacific fracture zones from satellite altimeter data. *Geophysical Journal* 69: 15–31.

Cazenave A, Lago B, and Dominh K (1983) Thermal parameters of the oceanic lithosphere estimated from geoid height data. *Journal of Geophysical Research* 88: 1105–1118.

Cazenave A, Parsons B, and Calcagno P (1995) Geoid lineations of 1000-km wavelength over the central Pacific. *Geophysical Research Letters* 22: 97–100.

Chase CG (1985) The geological significance of the geoid. *Annual Review of Earth and Planetary Sciences* 13: 97–117.

Chastel YB, Dawson PR, Wenk H-R, and Bennett K (1993) Anisotropic convection with implications for the upper mantle. *Journal of Geophysical Research* 98: 17757–17772. http://dx.doi.org/10.1029/93JB01161.

Choblet G and Parmentier EM (2001) Mantle upwelling and melting beneath slow spreading centers: Effects of variable rheology and melt productivity. *Earth and Planetary Science Letters* 184: 589–604.

Clouard V and Bonneville A (2001) How many Pacific hotspots are fed by deep-mantle plumes? *Geology* 29: 695–698.

Colin P and Fleitout L (1990) Topography of the ocean floor: Thermal evolution of the lithosphere and interaction of the deep mantle heterogeneities with the lithosphere. *Geophysical Research Letters* 17: 1961–1964.

Conrad CP, Benjun W, Smith EI, Bianco TA, and Tibbetts A (2010) Shear-driven upwelling induced by lateral viscosity variations and asthenospheric shear: A mechanism for intraplate volcanism. *Physics of the Earth and Planetary Interiors* 178(3): 162–175.

Conrad CP, Bianco TA, Smith EI, and Wessel P (2011) Patterns of intraplate volcanism controlled by asthenospheric shear. *Nature Geoscience* 4: 317–321.

Cooper RF (2002) Seismic wave attenuation: Energy dissipation in viscoelastic crystalline solids. In: Karato S and Wenk H-R (eds.) *Plastic Deformation of Minerals and Rocks*. Washington, DC: Mineralogical Society of America.

Craig CH and McKenzie DP (1986) The existence of a thin low-viscosity layer beneath the lithosphere. *Earth and Planetary Science Letters* 78: 420–426.

Crosby AG and McKenzie D (2009) An analysis of young ocean depth, gravity and global residual topography. *Geophysical Journal International* 178: 1198–1219.

Crosby AG, McKenzie D, and Sclater JG (2006) The relationship between depth, age and gravity in the oceans. *Geophysical Journal International* 166: 553–573.

Crough ST (1975) Thermal model of oceanic lithosphere. *Nature* 256: 388–390.

Crough ST (1977) Approximate solutions for the formation of the lithosphere. *Earth and Planetary Science Letters* 14: 365–377.

Crough ST (1978) Thermal origin of mid-plate hot-spot swells. *Geophysical Journal of the Royal Astronomical Society* 55: 451–469.

Crough ST (1979a) Geoid anomalies across fracture zones and the thickness of the lithosphere. *Earth and Planetary Science Letters* 44: 224–230.

Crough ST (1979b) Hotspot epeirogeny. *Tectonophysics* 61: 321–333.

Davaille A and Jaupart C (1994) One onset of thermal convection in fluids with temperature-dependent viscosity: Application to the ocean mantle. *Journal of Geophysical Research* 99: 19853–19866.

Davis EE and Lister CRB (1974) Fundamentals of ridge crest topography. *Earth and Planetary Science Letters* 21: 405–413.

Detrick RS (1981) An analysis of geoid anomalies across the Mendocino fracture zone: Implications for thermal models of the lithosphere. *Journal of Geophysical Research* 86: 11751–11762.

Detrick RS and Crough ST (1978) Island subsidence, hot spots and lithosphere thinning. *Journal of Geophysical Research* 83: 1236–1244.

Doin MP and Fleitout L (1996) Thermal evolution of the oceanic lithosphere: An alternative view. *Earth and Planetary Science Letters* 142: 121–136.

Doin M-P, Fleitout L, and McKenzie D (1996) Geoid anomalies and the structure of continental and oceanic lithospheres. *Journal of Geophysical Research* 101: 16119–16136. http://dx.doi.org/10.1029/96JB00640.

Driscoll ML and Parsons B (1988) Cooling of the oceanic lithosphere – Evidence from geoid anomalies across the Udintsev and Eltanin fracture zones. *Earth and Planetary Science Letters* 88: 289–307.

Dunbar J and Sandwell D (1988) A boudinage model for cross-grain lineations (abs.). *EOS, Transactions, American Geophysical Union* 69: 1429.

Eberle MA and Forsyth DW (1995) Regional viscosity variations, small-scale convection and the slope of the depth-age ½ curve. *Geophysical Research Letters* 22: 473–476.

Ekstrom G and Dziewonski AM (1998) The unique anisotropy of the Pacific upper mantle. *Nature* 394: 168–172.

Elthon D and Scarfe CM (1984) High-pressure phase equilibria of a high-magnesia basalt and the genesis of primary oceanic basalts. *American Mineralogist* 69: 1–15.

Evans RL, Hirth G, Baba K, Forsyth D, Chave A, and Mackie R (2005) Geophysical evidence from the MELT area for compositional controls on oceanic plates. *Nature* 437: 249–252.

Faul U and Jackson I (2005) The seismological signature of temperature and grain size variations in the upper mantle. *Earth and Planetary Science Letters* 234: 119–234.

Fischer KM, Ford HA, Abt DL, and Rychert CA (2010) The lithosphere–asthenosphere boundary. *Annual Review of Earth and Planetary Sciences* 38: 551–575.

Fleitiout L and Yuen DA (1984) Secondary convection and the growth of the oceanic lithosphere. *Physics of the Earth and Planetary Interiors* 36: 181–212.

Fleitout L and Moriceau C (1992) Short-wavelength geoid, bathymetry and convective pattern beneath the Pacific Ocean. *Geophysical Journal International* 110: 6–28.

Forsyth DW (1992) Geophysical constraints on mantle flow and melt generation beneath mid-ocean ridges. In: Morgan P, Blackman DN, and Sinton M (eds.) *Evolution of Mid Ocean Ridges*. *Geophysical Monograph Series*, vol. 71, pp. 1–66. Washington, DC: American Geophysical Union.

Forsyth DW and The MELT Seismic Team (1998) Imaging the deep seismic structure beneath a mid-ocean ridge: The MELT experiment. *Science* 280: 1215–1218.

Frank FC (1968) Two component model for convection in the Earth's upper mantle. *Nature* 220: 350–352.

Freedman AP and Parsons B (1990) Geoid anomalies over two South Atlantic fracture zones. *Earth and Planetary Science Letters* 100: 18–41.

Gaboret C, Forte AM, and Montagner JP (2003) The unique dynamics of the Pacific Hemisphere mantle and its signature on seismic anisotropy. *Earth and Planetary Science Letters* 208: 219–233.

Gaherty JB, Jordan TH, and Lee LS (1996) Seismic structure of the upper mantle in a central Pacific corridor. *Journal of Geophysical Research* 101: 22291–22309.

Gaherty JB, Kato M, and Jordan TH (1999) Seismological structure of the upper mantle: A regional comparison of seismic layering. *Physics of the Earth and Planetary Interiors* 110(1–2): 21–41. http://dx.doi.org/10.1016/S0031-9201(98)00132-0.

Gans KD, Wilson DS, and Macdonald KC (2003) Pacific Plate gravity lineaments: Diffuse extension or thermal contraction? *Geochemistry, Geophysics, Geosystems* 4(9): 1074. http://dx.doi.org/10.1029/2002GC000465.

Goodwillie AM and Parsons B (1992) Placing bounds on lithospheric deformation in the central Pacific Ocean. *Earth and Planetary Science Letters* 111: 123–139.

Gribb TT and Cooper RF (1998) Low-frequency shear wave attenuation in polycrystalline olivine: Grain boundary diffusion and the physical significance of the Andrade model for viscoelastic rheology. *Journal of Geophysical Research* 103: 27267–27279.

Harmon N, Forsyth DW, and Scheirer DS (2006) Analysis of gravity and topography in the GLIMPSE study region: Isostatic compensation and uplift of the Sojourn and Hotu Matua Ridge systems. *Journal of Geophysical Research* 111: B11406. http://dx.doi.org/10.1029/2005JB004071.

Harmon N, Forsyth DW, Weeraratne DS, Yang Y, and Webb SC (2011) Mantle heterogeneity and off axis volcanism on young Pacific lithosphere. *Earth and Planetary Science Letters* 311(3): 306–315.

Haxby WF and Turcotte DL (1978) Isostatic geoid anomalies. *Journal of Geophysical Research* 83: 5473–5478.

Haxby WF and Weissel JK (1986) Evidence for small scale convection for SEASAT altimeter data. *Journal of Geophysical Research* 91: 3507–3520.

Hayes DE (1988) Age-depth relationships and depth anomalies in the Southeast Indian Ocean and South Atlantic Ocean. *Journal of Geophysical Research* 93: 2937–2954.
Heestand RL and Crough ST (1981) The effect of hot spots on the oceanic age–depth relation. *Journal of Geophysical Research* 86: 6107–6114.
Hernlund JW and Tackley PJ (2003) Post-Laramide volcanism and upper mantle dynamics in the western US: Role of small-scale convection. *EOS, Transactions, American Geophysical Union* 84(46), Fall Meet. Suppl., Abstract V31F-07.
Hirth G, Evans RL, and Chave AD (2000) Comparison of continental and oceanic mantle electrical conductivity: Is the Archean lithosphere dry? *Geochemistry, Geophysics, Geosystems* 1. http://dx.doi.org/10.1029/2000GC000048.
Hirth G and Kohlstedt DL (1996) Water in the oceanic upper mantle: Implications for rheology, melt extraction, and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144: 93–108.
Huang J and Zhong S (2005) Sublithospheric small-scale convection and its implications for the residual topography at old ocean basins and the plate model. *Journal of Geophysical Research* 110: B05404. http://dx.doi.org/10.1029/2004JB003153.
Huang J, Zhong S, and van Hunen J (2003) Controls on sublithospheric small-scale convection. *Journal of Geophysical Research* 108: 2405. http://dx.doi.org/10.1029/2003JB002456.
Jackson I and Faul UH (2010) Grainsize-sensitive viscoelastic relaxation in olivine: Towards a robust laboratory-based model for seismological application. *Physics of the Earth and Planetary Interiors* 183: 151–163.
Jackson I, Fitz Gerald JD, Faul UH, and Tan BH (2002) Grain-size sensitive seismic-wave attenuation in polycrystalline olivine. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2002JB001225.
Jha K, Parmentier EM, and Sparks DW (1997) Buoyant mantle upwelling and crustal production at oceanic spreading centers: On-axis segmentation and off-axis melting. *Journal of Geophysical Research* 102: 11979–11989.
Johnson HP and Carlson RL (1992) Variation of sea floor depth with age: A test of model based on drilling results. *Geophysical Research Letters* 19: 1971–1974.
Kaminski É and Ribe NM (2001) A kinematic model for recrystallization and texture development in olivine polycrystals. *Earth and Planetary Science Letters* 189: 253–267.
Kaminski É and Ribe NM (2002) Timescales for the evolution of seismic anisotropy in mantle flow. *Geochemistry, Geophysics, Geosystems* 3: 1–17. http://dx.doi.org/10.1029/2001GC000222.
Karato S (1986) Does partial melting reduce the creep strength of the upper mantle. *Nature* 319: 309–310.
Karato S (2012) On the origin of the asthenosphere. *Earth and Planetary Science Letters* 321–322: 95–103. http://dx.doi.org/10.1016/j.epsl.2012.01.001.
Katz RF (2010) Porosity-driven convection and asymmetry beneath mid-ocean ridges. *Geochemistry, Geophysics, Geosystems* 11: Q0AC07. http://dx.doi.org/10.1029/2010GC003282.
Katzman R, Zhao L, and Jordan TH (1998) High-resolution, two-dimensional vertical tomography of the central Pacific mantle using ScS reverberations and frequency-dependent travel times. *Journal of Geophysical Research* 103: 17933–17972. http://dx.doi.org/10.1029/98JB00504.
Kawakatsu H, et al. (2009) Seismic evidence for sharp lithosphere–asthenosphere boundaries of oceanic plates. *Science* 324: 499–502.
Kelemen PB, Hirth G, Shimizu N, Spiegelman M, and Dick HJB (1997) A review of melt migration processes in the adiabatically upwelling mantle beneath oceanic spreading ridges. *Philosophical Transactions of the Royal Society A* 355: 282–318.
King SD and Anderson DL (1998) Edge-driven convection. *Earth and Planetary Science Letters* 160: 289–296.
Koppers AAP, Staudigel H, Pringle MS, and Wijbrans JR (2003) Short-lived and discontinuous intraplate volcanism in the South Pacific: Hot spots or extensional volcanism? *Geochemistry, Geophysics, Geosystems* 4: 1089. http://dx.doi.org/10.1029/2003GC000533.
Korenaga J (2007) Effective thermal expansivity of Maxwellian oceanic lithosphere. *Earth and Planetary Science Letters* 257: 343–349.
Korenaga J and Jordan TH (2002a) On 'steady-state' heat flow and the rheology of oceanic mantle. *Geophysical Research Letters* 29: 2056. http://dx.doi.org/10.1029/2002GL016085.
Korenaga J and Jordan TH (2002b) Onset of convection with temperature- and depth-dependent viscosity. *Geophysical Research Letters* 29(19): 1923. http://dx.doi.org/10.1029/2002GL015672.
Korenaga J and Jordan TH (2003a) Physics of multiscale convection in Earth's mantle: Onset of sublithospheric convection. *Journal of Geophysical Research* 108(B7): 2333. http://dx.doi.org/10.1029/2002JB001760.
Korenaga J and Jordan TH (2003b) Linear stability analysis of Richter rolls. *Geophysical Research Letters* 30(22): 2157. http://dx.doi.org/10.1029/2003GL018337.
Korenaga J and Jordan TH (2004) Physics of multiscale convection in Earth's mantle: Evolution of sublithospheric convection. *Journal of Geophysical Research* 109: B01405. http://dx.doi.org/10.1029/2003JB002464.
Korenaga T and Korenaga J (2008) Subsidence of normal oceanic lithosphere, apparent thermal expansivity, and seafloor flattening. *Earth and Planetary Science Letters* 268(1–2): 41–51. http://dx.doi.org/10.1016/j.epsl.2007.12.022.
Kumar P and Kawakatsu H (2011) Imaging the seismic lithosphere–asthenosphere boundary of the oceanic plate. *Geochemistry, Geophysics, Geosystems* 12: Q01006. http://dx.doi.org/10.1029/2010GC003358.
Langseth MG, LePichon X, and Ewing M (1966) Crustal structure of mid-ocean ridges. 5. Heat flow through the Atlantic Ocean and convection currents. *Journal of Geophysical Research* 71: 5321–5355.
Larson RL (1991) The latest pulse of the Earth – Evidence for a mid-Cretaceous superplume. *Geology* 19: 547–550.
Larson RL and Erba E (1999) Onset of mid-Cretaceous greenhouse in the Barremian–Aptian: Igneous events and the biological, sedimentary and geochemical responses. *Paleoceanography* 14: 663–678.
Lawrence JF and Shearer PM (2006) A global study of transition zone thickness using receiver functions. *Journal of Geophysical Research* 111: B06307. http://dx.doi.org/10.1029/2005JB003973.
Lekic V and Romanowicz B (2011) Inferring upper-mantle structure by full waveform tomography with the spectral element method. *Geophysical Journal International* 185(2): 799–831. http://dx.doi.org/10.1111/j.1365-246X.2011.04969.x.
Levitt DA and Sandwell DT (1995) Lithospheric bending at subduction zones based on depth soundings and satellite gravity. *Journal of Geophysical Research* 100: 379–400.
Li XQ, et al. (2004) Rejuvenation of the lithosphere by the Hawaiian plume. *Nature* 427(6977): 827–829. http://dx.doi.org/10.1038/nature02349.
Lin J and Parmentier EM (1985) Surface topography due to convection in a variable viscosity fluid: Application to short wavelength gravity anomalies in the central Pacific Ocean. *Geophysical Research Letters* 12: 357–360.
Lister CRB, Sclater JG, Davis EE, Villinger H, and Nagihara S (1990) Heat flow maintained in ocean basins of great age: Investigations in the north-equatorial west Pacific. *Geophysical Journal International* 102: 603–630.
Lizarralde D, Chave A, Hirth G, and Schultz A (1995) Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data. *Journal of Geophysical Research* 100: 17837–17854.
Lundstrom CC (2003) Uranium-series disequilibria in mid-ocean ridge basalts: Observations and models of basalt genesis. In: Bourdon B, Henderson GM, Lundstrom CC, and Turner SP (eds.) *Uranium-Series Geochemistry. Reviews in Mineralogy and Geochemistry*, vol. 52, pp. 175–214. Washington, DC: Mineralogical Society of America.
Lynch MA (1999) Linear ridge groups: Evidence for tensional cracking in the Pacific Plate. *Journal of Geophysical Research* 104: 29321–29333.
Maggi A, Debayle E, Priestly K, and Barruol G (2006a) Multimode surface waveform tomography of the Pacific Ocean: A close look at the lithospheric cooling signature. *Geophysical Journal International* 166: 1384–1397.
Maggi A, et al. (2006b) Multimode surface waveform tomography of the Pacific Ocean: A closer look at the lithospheric cooling signature. *Geophysical Journal International* 166(3): 1384–1397. http://dx.doi.org/10.1111/j.1365-246X.2006.03037.x.
Maggi A, et al. (2006c) Azimuthal anisotropy of the Pacific region. *Earth and Planetary Science Letters* 250: 53–71. http://dx.doi.org/10.1016/j.epsl.2006.07.010.
Marquart G, Schmelling H, and Braun A (1999) Small-scale instabilities below the cooling oceanic lithosphere. *Geophysical Journal International* 138: 655–666.
Marty JC and Casanave A (1989) Regional variations in subsidence rate of oceanic plates: A global analysis. *Earth and Planetary Science Letters* 94: 301–315.
Marty J and Cazenave CA (1988) Thermal evolution of the lithosphere beneath fracture zones inferred from geoid anomalies. *Geophysical Research Letters* 15: 593–596.
Marty J, Cazenave CA, and Lago B (1988) Geoid anomalies across Pacific fracture zones. *Geophysical Journal* 93: 1–23.
McKenzie DP (1967) Some remarks on heat flow and gravity anomalies. *Journal of Geophysical Research* 72: 6261–6273.
McKenzie DP and Sclater JG (1969) Heat flow in the eastern Pacific and sea-floor spreading. *Bulletin of Volcanology* 33: 101–118.
McNutt MK (1998) Superswells. *Reviews of Geophysics* 36: 211–244.
Montagner J-P and Tanimoto T (1991) Global upper mantle tomography of seismic wave velocities and anisotropies. *Journal of Geophysical Research* 96: 20337–20351.
Morgan WJ (1968) Rises, trenches, great faults, and crustal blocks. *Journal of Geophysical Research* 73: 1959.

Morgan WJ (1971) Convection plumes in the lower mantle. *Nature* 230: 42–43.
Morgan WJ (1972) Deep mantle convection plumes and plate motions. *American Association of Petroleum Geologists Bulletin* 56: 203–213.
Nagihara S, Lister CRB, and Sclater JG (1996) Reheating of old oceanic lithosphere: Deductions from observations. *Earth and Planetary Science Letters* 139: 91–104.
Nettles M and Dziewonski AM (2008) Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America. *Journal of Geophysical Research* 113: B02303. http://dx.doi.org/10.1029/2006JB004819.
Nishimura CE and Forsyth DW (1989) The anisotropic structure of the upper mantle in the Pacific. *Geophysical Journal* 96: 203–229.
O'Hara MJ (1965) Primary magmas and the origin of basalts. *Scottish Journal of Geology* 1: 19–40.
Oikawa M, Kaneda K, and Nishizawa A (2010) Seismic structures of the 154–160 Ma oceanic crust and upper mantle in the Northwest Pacific Basin. *Earth, Planets and Space* 62: e13–e16.
Okal EA and Batiza R (1987) Hotspots: The first 25 years. In: Keating BH, Fryer P, Batiza R, and Boehlert GW (eds.) *Seamounts, Islands, and Atolls*. *Geophysical Monograph*, vol. 43, pp. 1–11. Washington, DC: AGU.
Parmentier EM and Haxby WF (1986) Thermal stresses in the oceanic lithosphere: Evidence from geoid anomalies at fracture zones. *Journal of Geophysical Research* 91: 7193–7204.
Parmentier EM and Phipps Morgan J (1990) The spreading rate dependence of three-dimensional oceanic spreading center structure. *Nature* 348: 325–328.
Parsons B and McKenzie D (1978) Mantle convection and the thermal structure of the plates. *Journal of Geophysical Research* 83: 4485–4496.
Parsons B and Sclater JG (1977) An analysis of the variation of ocean floor bathymetry and heat flow with age. *Journal of Geophysical Research* 82: 803–827.
Passey QR (1981) Upper mantle viscosity derived from the difference in rebound of the Provo and Bonneville shorelines: Lake Bonneville Basin, Utah. *Journal of Geophysical Research* 86: 11701–11708.
Phipps Morgan J (1987) Melt migration beneath mid-ocean spreading centers. *Geophysical Research Letters* 14: 1238–1241.
Phipps Morgan J (1997) The generation of a compositional lithosphere by mid-ocean ridge melting and its effect on subsequent off-axis hotspot upwelling and melting. *Earth and Planetary Science Letters* 146: 213–232.
Phipps Morgan J and Parmentier EM (1995) Crenulated seafloor: Evidence for spreading rate dependent structure of mantle upwelling and melting beneath a mid-oceanic spreading center. *Earth and Planetary Science Letters* 129: 73–84.
Priestley D and McKenzie DM (2006) The thermal structure of the lithosphere from shear velocities. *Earth and Planetary Science Letters* 244: 285–301.
Rabinowicz M and Ceuleneer G (2005) The effect of sloped isotherms on melt flow in the shallow mantle: A physical and numerical model based on observations in the Oman ophiolite. *Earth and Planetary Science Letters* 229: 231–246.
Raddick MJ, Parmentier EM, and Scheirer DS (2002) Buoyant decompression melting: A possible mechanism for intraplate volcanism. *Journal of Geophysical Research* 107(B10): 2228. http://dx.doi.org/10.1029/2001JB000617.
Renkin ML and Sclater JG (1988) Depth and age in the North Pacific. *Journal of Geophysical Research* 93: 2910–2935.
Richards MA, Duncan RA, and Courtillot VE (1989) Flood basalts and hot-spot tracks: Plume heads and tails. *Science* 246: 103–107.
Richter FM (1973) Convection and the large scale circulation of the mantle. *Journal of Geophysical Research* 74: 8735–8745.
Richter FM and Parsons B (1975) On the interaction of two scales of convection in the mantle. *Journal of Geophysical Research* 80: 2529–2541.
Ritzwoller MH, Shapiro NM, and Zhong SJ (2004) Cooling history of the Pacific lithosphere. *Earth and Planetary Science Letters* 226(1–2): 69–84. http://dx.doi.org/10.1016/j.epsl.2004.07.032.
Rychert CA, Schmerr N, and Harmon N (2012) The Pacific lithosphere–asthenosphere boundary: Seismic imaging and anisotropic constraints from SS waveforms. *Geochemistry, Geophysics, Geosystems* 13: Q0AK10. http://dx.doi.org/10.1029/2012GC004194.
Rychert CA and Shearer PM (2011) Imaging the lithosphere–asthenosphere boundary beneath the Pacific using SS waveform modeling. *Journal of Geophysical Research* 116: B07307. http://dx.doi.org/10.1029/2010JB008070.
Sandwell DT (1984) Thermomechanical evolution of oceanic fracture zones. *Journal of Geophysical Research* 89: 11401–11413.
Sandwell D and Fialko Y (2004) Warping and cracking of the Pacific plate by thermal contraction. *Journal of Geophysical Research* 109: B10411. http://dx.doi.org/10.1029/2004JB003091.
Sandwell DT and McAdoo DC (1988) Marine gravity of the southern ocean and Antarctic margin from Geosat. *Journal of Geophysical Research* 93: 10389–10396.
Sandwell D and Schubert G (1980) Geoid height versus age for symmetric spreading ridges. *Journal of Geophysical Research* 85: 7235–7241.
Sandwell DT and Schubert G (1982a) Geoid height-age relation from SEASAT altimeter profiles across the Mendocino fracture zone. *Journal of Geophysical Research* 87: 3949–3958.
Sandwell DT and Schubert G (1982b) Lithospheric flexure at fracture zones. *Journal of Geophysical Research* 87: 4657–4667.
Sandwell DT and Smith WHF (1997) Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *Journal of Geophysical Research* 102: 10039–10054. http://dx.doi.org/10.1029/96JB03223.
Sandwell DT, Winterer EL, Mammerickx J, et al. (1995) Evidence for diffuse extension of the Pacific plate from Pukapuka Ridges and Cross-Grain gravity lineations. *Journal of Geophysical Research* 100: 15087–15099.
Schlanger SO, Garcia MO, Keating BH, et al. (1984) Geology and geochronology of the Line Islands. *Journal of Geophysical Research* 89: 11261–11272.
Schmerr N (2012) The Gutenberg discontinuity: Melt at the lithosphere–asthenosphere boundary. *Science* 335(6075): 1480–1483. http://dx.doi.org/10.1126/science.1215433.
Schroeder W (1984) The empirical age–depth relation and depth anomalies in the Pacific Ocean. *Journal of Geophysical Research* 89: 9873–9884.
Schubert G, Froidevaux C, and Yuen DA (1976) Oceanic lithosphere and asthenosphere: Thermal and mechanical structure. *Journal of Geophysical Research* 81: 3525–3540.
Schubert G and Turcotte DL (1972) One dimensional model of shallow mantle convection. *Journal of Geophysical Research* 77: 945–951.
Schubert G, Yuen DA, Froidevaux C, Fleitout L, and Souriau M (1978) Mantle circulation with partial shallow return flow. *Journal of Geophysical Research* 83: 745–758.
Sclater JG, Lawver LA, and Parsons B (1975) Comparison of long-wavelength residual elevation and free air gravity anomalies in the north Atlantic and possible implications for thickness of the lithospheric plate. *Journal of Geophysical Research* 80: 1031–1052.
Sigmundsson F (1991) Post-glacial rebound and asthenosphere viscosity in Iceland. *Geophysical Research Letters* 18: 1131–1134. http://dx.doi.org/10.1029/91GL01342.
Sleep NH (1969) Sensitivity of heat flow and gravity to the mechanism of seafloor spreading. *Journal of Geophysical Research* 74: 542–549.
Smith WHF (1998) Seafloor tectonic fabric from satellite altimetry. *Annual Review of Earth and Planetary Sciences* 26: 697–738.
Sparks DW and Parmentier EM (1991) Melt extraction from the mantle beneath spreading centers. *Earth and Planetary Science Letters* 105: 368–377.
Sparks DW and Parmentier EM (1993) The structure of three-dimensional convection beneath spreading centers. *Geophysical Journal International* 112: 81–91.
Spiegelman M (1993) Physics of melt extraction: Theory, implications and applications. *Philosophical Transactions of the Royal Society A* 342: 23–41.
Spiegelman M and Elliot T (1993) Consequences of melt transport for Uranium series disequilibrium. *Earth and Planetary Science Letters* 118: 1–20.
Spiegelman M and Kelemen PB (2003) Extreme chemical variability as a consequence of channelized melt transport. *Geochemistry, Geophysics, Geosystems* 4(7): 1055. http://dx.doi.org/10.1029/2002GC000336.
Spiegelman M and McKenzie D (1987) Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth and Planetary Science Letters* 83: 137–152.
Stein CA and Stein S (1992) A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* 359: 123–129.
Stein CA and Stein S (1993) Constraints on Pacific midplate swells from global depth-age and heat flow-age models. In: Pringle MS, et al. (eds.) *The Mesozoic Pacific: Geology, Tectonics, and Volcanism*. *Geophysical Monograph Series*, vol. 77, pp. 53–76. Washington, DC: AGU.
Stixrude L and Lithgow-Bertelloni C (2005) Mineralogy and elasticity of the oceanic upper mantle: Origin of the low-velocity zone. *Journal of Geophysical Research* 110: B03204. http://dx.doi.org/10.1029/2004JB002965.
Stolper E (1980) A phase diagram for mid-ocean ridge basalts: Preliminary results and implications for petrogenesis. *Contributions to Mineralogy and Petrology* 74: 13–27.
Sundberg M and Cooper RF (2010) A composite viscoelastic model for incorporating grain boundary sliding and transient diffusion creep: Correlating creep and attenuation responses for materials with a fine grain size. *Philosophical Magazine* 90: 2817–2840.
Tackley PJ and Stevenson DJ (1993) A mechanism for spontaneous self-perpetuating volcanism on the terrestrial planets. In: Stone DB and Runcorn SK (eds.) *Flow and Creep in the Solar System: Observations, Modeling and Theory*, pp. 307–321. Norwell, MA: Kluwer Academic.
Tan Y and Helmberger DV (2007) Trans-Pacific upper mantle shear velocity structure. *Journal of Geophysical Research* 112: B08301. http://dx.doi.org/10.1029/2006JB004853.

Turcotte DL and Oxburgh ER (1967) Finite amplitude convective cells and continental drift. *Journal of Fluid Mechanics* 28: 24–42.

Turner DL and Jarrard RD (1982) K–Ar dating of the Cook–Austral island chain: A test of the hot-spot hypothesis. *Journal of Volcanology and Geothermal Research* 12: 187–220.

van Hunen J, Huang J, and Zhong S (2003) The effect of shearing on the onset and vigor of small-scale convection in a Newtonian rheology. *Geophysical Research Letters* 30(19): 1991. http://dx.doi.org/10.1029/2003GL018101.

Vogt PR and Ostenso NA (1967) Steady state crustal spreading. *Nature* 215: 811–817.

Wang D, Mookerjee M, Xu Y, and Karato S (2006) The effect of water on the electrical conductivity of olivine. *Nature* 443: 977–980. http://dx.doi.org/10.1038/nature05256.

Weeraratne DS, Forsyth DW, Yang Y, and Webb SC (2007) Rayleigh wave tomography beneath intraplate volcanic ridges in the South Pacific. *Journal of Geophysical Research* 112(B6): B06303.

Wenk H-R and Tomé CN (1999) Modeling dynamic recrystallization of olivine aggregates deformed in simple shear. *Journal of Geophysical Research* 104: 25513–25528. http://dx.doi.org/10.1029/1999JB900261.

Wessel P (2001) Global distribution of seamounts inferred from gridded Geosat/ERS-1 altimetry. *Journal of Geophysical Research* 106: 19431–19441.

Wessel P and Haxby WF (1990) Thermal-stresses, differential subsidence, and flexure at oceanic fracture zones. *Journal of Geophysical Research* 95: 375–391.

Wessel P, Kroenke LW, and Bercovici D (1996) Pacific Plate motion and undulations in geoid and bathymetry. *Earth and Planetary Science Letters* 140: 53–66.

Wessel P and Lyons S (1997) Distribution of large Pacific seamounts from Geosat/ERS-1: Implications for the history of intraplate volcanism. *Journal of Geophysical Research* 102: 22459–22475.

Yoshino T, Matsuzaki T, Yamashita S, and Katsura T (2006) Hydrous olivine unable to account for conductivity anomaly at the top of the asthenosphere. *Nature* 443: 973–976. http://dx.doi.org/10.1038/nature05223.

Yuen DA and Fleitout L (1984) Stability of the oceanic lithosphere with variable viscosity: An initial-value approach. *Physics of the Earth and Planetary Interiors* 343: 173–185.

Zaranek SE and Parmentier EM (2004) Convective instability of a fluid with temperature-dependent viscosity cooled from above. *Earth and Planetary Science Letters* 224: 371–386.

Zhou Y, Nolet G, Dahlen FA, and Laske G (2006) Global upper-mantle structure from finite-frequency surface-wave tomography. *Journal of Geophysical Research* 111: B04304. http://dx.doi.org/10.1029/2005JB003677.

This page intentionally left blank

7.09 Dynamics of Subducting Slabs: Numerical Modeling and Constraints from Seismology, Geoid, Topography, Geochemistry, and Petrology

I Wada, Tohoku University, Sendai, Miyagi, Japan
S King, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by S D King, Volume 7, pp. 325–370, © 2007, Elsevier B.V.

7.09.1	**Introduction**	339
7.09.2	**Observations on Slab Dynamics**	340
7.09.2.1	Slab Geometry Inferred from Seismological Observations	341
7.09.2.2	Petrologic Structures of Subducting Slabs Inferred from Seismic Data	343
7.09.2.3	Slab Dehydration and Melting: Geochemistry of Arc Lava and the Location of the Arc	345
7.09.2.4	Subduction Thrust Earthquakes, Episodic Tremor and Slip, and Intraslab Earthquakes	346
7.09.2.5	Shear-Wave Splitting in Subduction Zones	349
7.09.2.6	Correlations among Slab Buoyancy, Kinematics, and Geometry	350
7.09.2.7	Geoid and Topography	353
7.09.3	**Kinematic Models for Slab Thermal Structure**	355
7.09.3.1	McKenzie Model	356
7.09.3.2	Corner Flow Solution	357
7.09.3.3	Earlier Estimates of Frictional Heating Along the Plate Interface	358
7.09.4	**Coupled Kinematic–Dynamic Slab–Wedge Models**	359
7.09.4.1	Governing Equations and Boundary Conditions	360
7.09.4.2	Slab–Mantle Viscous Coupling	361
7.09.4.3	Isoviscous to Temperature-Dependent Mantle Rheologies	362
7.09.4.4	Slab Surface and Subarc Mantle Temperatures	363
7.09.4.5	Thermopetrologic Models: Fluid Availability and Subduction Zone Processes	365
7.09.4.6	Effects of 3-D Slab Geometry on Thermal Structure	367
7.09.5	**Dynamic Subduction Models**	368
7.09.5.1	Simplifying Approximations from Compressible to Incompressible	368
7.09.5.2	From Mantle Downwelling to Plate Subduction	369
7.09.5.3	Effects of Phase Transformations and Thermodynamic Properties on Slab Dynamics	370
7.09.5.4	Role of Mantle Rheology and Slab Strength	374
7.09.5.5	Slab Rollback and Shallow-Dipping Slab	375
7.09.5.6	Plate Bending	377
7.09.6	**Instantaneous Mantle Flow Models: Geoid and Plate Motion**	378
7.09.7	**Concluding Remarks**	379
Acknowledgments		381
References		381

7.09.1 Introduction

Subduction zones are a major component of the coupled system of plate tectonics and mantle convection and in particular form the downwelling limb of the mantle's convective cold thermal boundary layer at the Earth's surface. Subducting slabs are the dominant source of gravitational potential energy (GPE) release that is the driving force for mantle convection and plate tectonics. It is through subducted slabs that oceanic crust and sediments, including volatile compounds such as water, are recycled back into the mantle. The Earth is the only terrestrial planet with active subduction zones, and it is possible that it is the only planet where the process of subduction as we know it ever occurred although there are some possible sites of past subduction processes on Venus (e.g., Schubert and Sandwell, 1995). Without the recycling of volatile compounds back into the mantle at subduction zones, it is possible that the Earth would have developed a runaway greenhouse similar to Venus. It is conceivable that without subduction, the Earth would not have been a suitable place for life to develop.

One of the most important driving forces of subduction is the negative thermal buoyancy of the subducting slabs, referred to as slab pull. Cooling of newly formed, buoyant oceanic plates near mid-ocean ridges causes the plates to subside and slide away from the ridge due to increasing density, and this ridge push also contributes to the net driving force. As a plate subducts, temperature- and pressure-dependent phase transformations and metastability of minerals in the slab, particularly that of olivine, influence the slab's negative buoyancy relatively to the surrounding mantle, possibly altering the path of the slab. Further, the driving force must be sufficiently large to overcome the frictional force exerted by the overriding

plate and the bending force required to cause the change in the slab dip as it subducts. The slab at depth also experiences viscous resistance exerted by the surrounding mantle, and the slab may become even anchored in highly viscous mantle. Quantifying these forces and understanding how subducting slab evolve thermally and petrologically during subduction are the critical part of and the greatest challenges in the studies of subduction dynamics.

There have been a number of reviews on the studies of the structure and dynamics of subduction zones (e.g., Bebout et al., 1996; Becker and Faccenna, 2009; Billen, 2008; Eiler, 2003; Gerya, 2011; King, 2001; Stern, 2002; van Keken and King, 2005) with some specifically focusing on seismology (Lay, 1994) the mantle wedge (van Keken, 2003; Wiens et al., 2008), phase changes and the fate of slabs (Christensen, 1995), slab stagnation in the transition zone (Fukao et al., 2009), and water in the mantle (Hirschmann, 2006). In an attempt to provide more than a summary of these reviews and recent papers, we will focus this chapter on vital themes that emerge from recent observations and modeling of plate subduction and mantle convection.

Our understanding of the dynamics of subducting slabs comes from a range of observations, including seismological observations, gravity, topography, surface heat flow, and geochemistry and petrology of island arc lavas, as well as from analog and numerical models constrained by these observations. Seismological observations, such as relocation of earthquake hypocenters, seismic tomography, and shear-wave splitting, provide crucial information about the geometry, composition, and structure of the subducting slab. These observations indicate variable behavior and geometry of subducting slabs among different subduction zones, and they also indicate significant 3-D structure and complexity within subduction zones (see **Chapter 1.20**). Gravity and topography are interrelated observations because elevation differences give rise to gravitational anomalies. Gravity and gravitational potential (or geoid) anomalies and along with topography observations provide a constraint on mantle rheology and the relative strength of the subducting slab. At length scales on the order of 100 km, they also provide information about the history of volcanic construction and mass transport due to erosion. The downward advection of a relatively cold oceanic lithosphere cools the surrounding mantle, yet mantle wedge flow driven by the motion of the slab brings hot material into the mantle wedge and provides the thermal condition required for melt generation and arc volcanism. Surface heat flow observations help us understand the fine interplay of the subducting slab and the hot overlying mantle wedge. Geochemical and petrologic data of arc lavas provide information about the geochemical flux from the subducting slabs, and experimental petrology allows us to estimate the metamorphic conditions and melting conditions in the slab and the overlying mantle. In many cases, the observables, such as slab dip, convergence rate, and arc geochemistry, show little correlation with one another, reflecting the complexity of subduction system. In **Section 7.09.2**, we discuss in more detail what these observations tell us about the dynamics of slabs and their interactions with the surrounding mantle.

Complementary to the increasing volume and complexity of the observations, various numerical modeling approaches have been developed to further understanding of the thermal structure and the dynamics of subducting slabs and mantle convection. In the early days of numerical subduction modeling, a common approach to study the thermal structure of the subducting slab was to use kinematic models. In this approach, the velocity field of the slab and mantle wedge was predefined, and the resulting thermal structure was calculated. The use of kinematic models has declined significantly since the development of coupled kinematic–dynamic models for a system of a subducting slab and overlying mantle wedge. In the coupled model, the motion of the subducting slab is kinematically described while the flow of the mantle in the overlying mantle wedge is solved, and the thermal field is solved for the entire model domain. In dynamic subduction zone models, the thermal (and sometimes chemical) buoyancy of the slab drives the motion of the slab and the larger-scale mantle flow, allowing for simulation of the evolution of the subducting slab and the surrounding mantle. The resolution and the complexity of both dynamic and kinematic–dynamic models are increasing, and some models include the effects of melting, phase changes, dehydration, and variable rheology. These approaches have recently been applied to three-dimensional geometries, owing to the advances in numerical methods and computational capability. To better understand the deformation field within the slab in regional-scale dynamic models, viscoelastic or viscoelastic–plastic rheology is used. Increasingly, regional-scale models in 2-D and 3-D are being applied to a specific subduction system in order to explain a range of observations (e.g., shear-wave splitting, heat flow, and volcanic history) that could not be treated with earlier generations of models. Instantaneous models solve for the deformation field that is in equilibrium with the boundary conditions and predefined loads and are used in conjunction with observables (e.g., surface plate motion, and geoid) to constrain unknown fields, such as viscosity and density. In **Sections 7.09.3** and **7.09.6**, we discuss the key characteristics and applications of kinematic, kinematic–dynamic, dynamic, and instantaneous models, respectively. The differences in the model types and approaches are also discussed by Billen (2008), and recent studies of numerical subduction modeling are summarized in Gerya (2011).

There have been many advances in all these areas in the past decade, and it would be impossible to include all the important contributions that have been made. We will defer to other review papers that highlight areas to which we cannot give proper attention.

7.09.2 Observations on Slab Dynamics

The structure and geometry of subducting slabs have been inferred from a range of seismological observations, including seismic refraction and reflection data, seismic-wave dispersion and scattering, intraslab earthquake hypocenters, and seismic tomography. We can also infer the state of stress in the subducting slab from intraslab earthquake focal mechanisms and moment tensor solutions. Shear-wave splitting data further provide us with the orientations of micro- and macro-structures, such as mineral fabrics, fractures, melt lenses, and faults. Geoid observations also tell us about the interaction of

subducting slabs with the surrounding mantle. Petrologic data on the subducting material (input) and the geochemistry of arc lavas (output) provide critical information about the role of subducting slabs on global volatile cycles, arc volcanism, and the composition and rheology of the Earth's mantle. In this section, we briefly summarize these observations that form the foundation of our understanding of the evolution and dynamics of subducting slabs. These observations are the key constraints for a number of numerical and laboratory modeling studies, which are discussed in **Sections 7.09.4–7.09.7**.

7.09.2.1 Slab Geometry Inferred from Seismological Observations

The geometry of subducted slabs was first constrained by earthquake hypocenters and focal mechanisms (e.g., Isacks and Molnar, 1969; Isacks et al., 1968), and the development of the worldwide seismograph network enabled the accurate location of earthquake hypocenters. Gutenberg and Richter (1939) first classified earthquakes by focal depth and compiled the distribution of global earthquake hypocenters clearly showing that intermediate and deep-focus earthquakes occurred in narrow zones that are consistent with descending plate material beneath trenches (cf. Isacks and Molnar, 1969). The 2-D slab surface geometry in the forearc and arc regions has been delineated from the distribution of earthquake hypocenters globally (e.g., Bailey et al., 2012; Engdahl et al., 1995; Syracuse and Abers, 2006).

Through the integration of seismological observations, detailed 3-D models of the slab surface geometry in the forearc and arc regions have been constructed for some subduction zones; one excellent example is the slab surface model of the Pacific slab in NE Japan where there are dense seismograph networks and high level of seismicity, providing a vast quantity of seismic data (e.g., Kita et al., 2010; Nakajima and Hasegawa, 2006; Zhao et al., 1997) (**Figure 1(a)**). In other subduction zones, slab geometry in the forearc and arc regions is constrained by a compilation of a range of seismological observations. Wide-angle seismic reflection and refraction are commonly used to image the structure of the shallow portion (<~50 km depth) of subduction zones. For the deeper part, seismic tomography is a common approach. For example, McCrory et al. (2012, references therein) constructed a geometric model for the subducting Juan de Fuca slab in the entire Cascadia subduction zone by compiling published seismic reflection and refraction data for the shallow part of the slab and intraslab earthquake hypocenters and teleseismic tomographic images of seismic velocity structures for the deeper part (**Figure 1(b)**). These models show local- to regional-scale along-strike variations in the slab geometry, illustrating the complexity of the slab geometry. Curved margins, such as in the northern part of NE Japan and Cascadia (**Figure 1**), result in an along-strike change in the dip direction of the slab and regional upward or downward warping of the slab. A large change in the slab dip along the strike of the margin also occurs in some subduction zones, such as in Central America, where the dip angle changes from about 80° in Nicaragua with intraslab earthquakes occurring to about ~200 km depth to 60° in the central Costa Rica with earthquakes occurring to ~125 km depth, and even shallower (dip and maximum earthquakes) to the south (Colombo et al., 1997). Imaging the shape of the slab is a critical part of the study of subduction geodynamics as

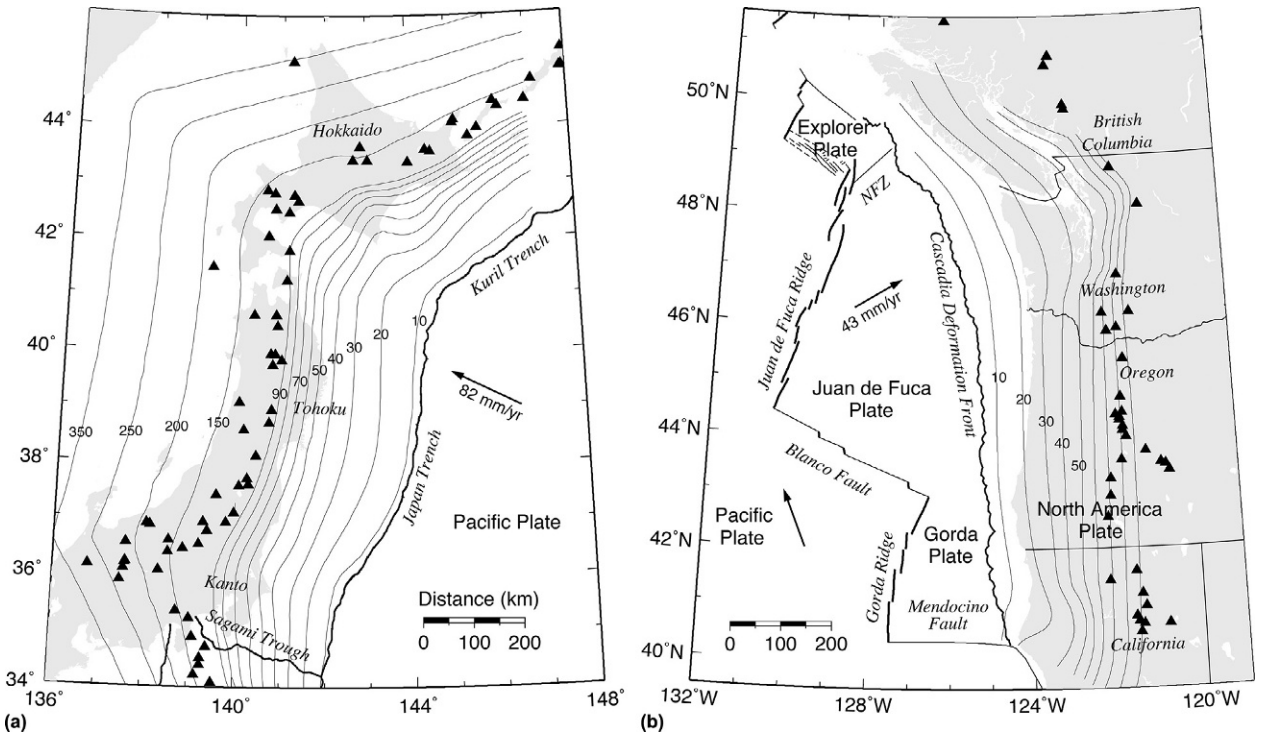


Figure 1 3-D slab geometry model for (a) NE Japan from Kita et al. (2010) and Nakajima and Hasegawa (2006) and (b) Cascadia from McCrory et al. (2012). Thin lines in (a) and (b) indicate the slab surface depth contours (km), and triangles indicate the locations of volcanoes.

slab geometry is influenced by tectonic forces acting on the slab, the rheology of the slab, the thermal and petrologic structures of the slab, and the dynamics of plate subduction. They are also a critical part of constructing geophysical models, such as coupled kinematic–dynamic thermal models, dislocation models for interseismic deformation, and tsunami models.

At larger scales, seismic tomography has become the primary tool for imaging the subducting slab, particularly in the lower mantle where intraslab seismicity is absent. Subducting slabs are imaged as long, often narrow traces of seismically fast materials. With an increasing number of tomographic studies in subduction zones, it is becoming clear that slabs develop a complex geometry as they subduct, and the behavior of slabs varies not only among different subduction zones but also within the subduction zones (Bunge and Grand, 2000; Chiu et al., 1991; Fukao et al., 1992, 2009; Gao et al., 2006; Grand, 2002; Gudmundsson and Sambridge, 1998; Li et al., 2008; Miller et al., 2005, 2004; Simmons et al., 2012; Tajima and Grand, 1998; van der Hilst, 1995; van der Hilst et al., 1993, 1991; van der Hilst and Seno, 1993; Widiyantoro et al., 1999; Yang et al., 2009; Zhou and Clayton, 1990).

A summary of tomographic cross sections across the Pacific subduction zones (**Figure 2**) from Karason (2002) illustrates both a variety of slab planforms; in Izu–Bonin and Tonga, the slabs lie flat on the base of the transition zone, while in Indonesia, central Japan, and Central America, the slabs traverse the transition with little apparent resistance, and some slabs appear to thicken in the lower mantle. Beneath northern Tonga, the flat-lying slab extends for almost 1000 km before descending into the lower mantle (van der Hilst, 1995). Flat-lying fast seismic velocity anomalies at the base of the transition zone have been found beneath northern Japan (van der Hilst et al., 1991, 1993). These have been investigated via dynamic subduction modeling, which is discussed in **Section 7.09.5**. In the upper mantle, the slab appears to become thinner with depth. It is possible that the general appearance of thinning is a result of the seismic properties of the mantle in this region, rather than reflecting a thinning of the slab itself. Without independently solving for the topography of phase boundaries, travel-time anomalies due to the deflection of the phase boundary will be mapped into volume perturbations in the 3-D seismic model and cause a jump in velocity on the order of 5%, a significant contribution to the total travel-time anomaly. If the horizontally averaged velocity differs from reference model at a given depth, this will lead to the apparent thickening or thinning of vertical anomalies when plotting cross sections of perturbations relative to a reference model. Thus, especially near phase boundaries, caution should be exercised when interpreting seismic cross sections. Li et al. (2008) updated the global P-wave tomography model of Karason and van der Hilst (2001), incorporating new sets of data and utilizing a wider range of seismic phases to maximize the sampling of the Earth's interior and using new approaches such as an irregular parameterization to increase resolution regionally. Their model shows the structural complexity of subducting slabs even more clearly. In addition to the slab planforms inferred from Karason and van der Hilst (2001), the updated model shows that the slab sinks directly into the lower mantle beneath Kamchatka, the Marianas, and

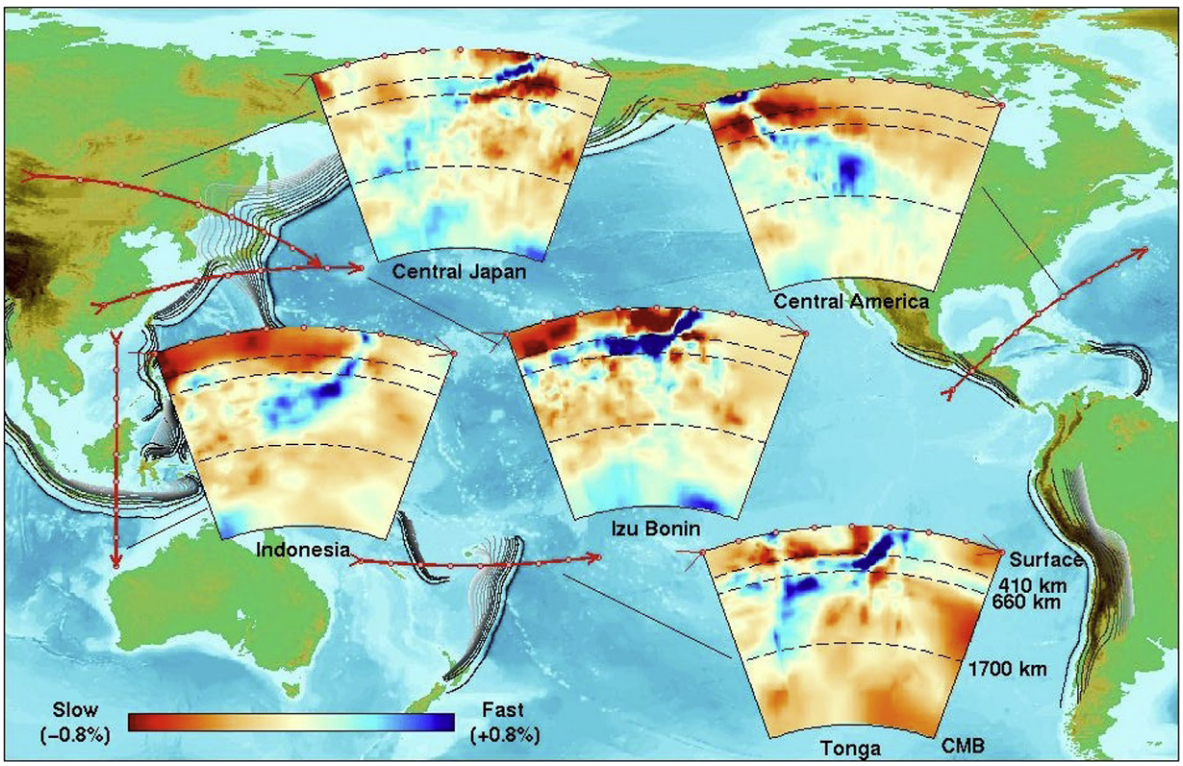


Figure 2 Summary of seismic tomography P-wave cross sections through Pacific subduction zones from Karason (2002). Reproduced from Karason H (2002) *Constraints on Mantle Convection from Seismic Tomography and Flow Modeling*. PhD Thesis, MIT.

the Celebes Sea. The flat-lying slabs in the transition zone, particularly along western Pacific, are clearly imaged in the more recent global P-wave model by Simmons et al. (2012). However, as Simmons et al. (2012) pointed out, the amplitudes and lateral extent of the fast anomalies that are interpreted as flat-lying slabs in the transition zone differ among different tomography models due to differences in the modeling approaches.

The stagnation of the subducting slab in the transition zone has drawn significant attention in the subduction geodynamics community. Fukao et al. (2009) provided a comprehensive overview of the past observational studies on slab stagnation. The effects of various factors, such as the mantle viscosity structure and trench motion, on slab stagnation have been investigated via dynamic modeling and are discussed in detail in **Section 7.09.5**. In Northeast Asia, the flat slab in the transition zone has been linked to the interplate volcanism (Fukao et al., 2001; Zhao et al., 2004, 2009). This link has led to the idea of the 'Big Mantle Wedge,' in which hot wet asthenospheric mantle material ascends above the flat-lying slab, generating volcanism (Zhao et al., 2009). A recent numerical modeling study (Kameyama and Nishioka, 2012) that investigates the role of the flat-lying slab in the 'Big Mantle Wedge' model will be discussed in **Section 7.09.5**. The amount of apparent thickening of the slab in the lower mantle cannot be explained by simple compressive thickening (e.g., Ribe et al., 2007). One possible mechanism for the apparent thickening is slab buckling, and the dynamic modeling results for slab buckling are also discussed in **Section 7.09.5**.

The slab geometry is further complicated by local anomalies. For example, analyses of tomographic slab models (Miller et al., 2005; Obayashi et al., 2006, 2009) show a distinct change in the seismic velocity in the Pacific slab beneath the Izu–Bonin arc at 300–450 km depth, in a position north and west of the Ogasawara Plateau at the southern end of the Izu–Bonin arc. The change in morphology of the slab could be interpreted as a 'slab tear.' In Tonga, there is a group of well-located deep earthquakes that lies about 200 km above the deepest end of the Tonga slab (Brudzinski and Chen, 2003a,b; Chen and Brudzinski, 2001, 2003; Fischer et al., 1991; Hamburger and Isacks, 1987; van der Hilst, 1995) (see **Chapters 4.13** and **4.18**). If this cluster of earthquakes represents a remnant piece of the Tonga slab, a possible explanation for the origin of a detached slab is the subduction of the Pacific Plate along the fossil Vityaz Trench where subduction ceased about 5–8 million years ago (Chen and Brudzinski, 2001; Fischer et al., 1991; Hamburger and Isacks, 1987; Okal and Kirby, 1998). Another interpretation is that this cluster of earthquakes represents a subhorizontal extension of the Tonga slab along the subduction zone (Fukao et al., 1992; van der Hilst et al., 1993). The interaction of the Tonga slab with a mantle upwelling (Gurnis et al., 2000a) is likely to be a significant factor (perhaps in combination with the other events). With either explanation, it is clear that the interaction of subducted material with the transition zone can yield complex planforms of subducted material. The gap in seismicity between roughly 300 and 600 km depths in the South American slab has been observed since the earliest studies of subduction zones (Isacks and Molnar, 1971). Receiver function analysis of the 410-km discontinuity along the length of the South American subduction zone is consistent with the presence of a continuous slab through the aseismic region (Collier et al., 2001). The shape of the 660 km discontinuity from receiver functions (Collier et al., 2001) and the presence of lower mantle reflectivity in shear waves (Contenti et al., 2012) are consistent with both stagnation and penetration as the Nazca slab impinges upon the base of the upper mantle. However, a more recent study of the reconstruction of the motion of the Nazca and South American Plates that includes a change in the direction of the Farallon Plate at 25 Ma predicts the presence of a slab window beneath South America due to the subduction of the Farallon–Phoenix spreading ridge, consistent with the aseismic region of the South American slab (O'Driscoll et al., 2012). These studies are examples of the growing number of observations of the complexity of the slab geometry.

7.09.2.2 Petrologic Structures of Subducting Slabs Inferred from Seismic Data

A subducting slab is composed of layers of compositionally distinct components: sediments, oceanic crust, and oceanic mantle. The oceanic mantle makes up more than 90% of the mass (volume) of a subducting slab, and its buoyancy and rheology largely control the dynamics of subduction. The mass (volume) of oceanic crust and sediment is relatively small, but they play key roles in geochemical cycles and also control the mechanical behavior of the interface between the slab and the overlying mantle.

The fresh oceanic crust that is created at the mid-ocean ridge, the mid-ocean ridge basalt (MORB), contains little water and is depleted in incompatible elements (i.e., those that preferentially partition into the melt phase) such as potassium (K) and uranium (U), due to repeated episodes of partial melting of the magma source (e.g., Dixon et al., 2002). However, the topmost portion of the oceanic crust is relatively porous and permeable due to lava drainage, fissuring, and fractures, promoting pervasive hydrothermal circulation and alteration (Davis et al., 1997), resulting in the enrichment of MORB with some incompatible elements (e.g., Staudigel et al., 1996; Wallmann, 2001). Hydrothermal alteration also results in the formation of hydrous minerals, such as chlorite, epidote/zoisite, amphibole (hornblende, actinolite, and tremolite), chloritoid, talc, and phengite (Schmidt and Poli, 1998). The lower portion of the crust is likely to be hydrated locally along faults except under unusual conditions.

Oceanic mantle lithosphere is peridotitic in composition (mostly olivine and pyroxene), and its upper portion consists of the residual (harzburgite) of mantle material after MORB formation, whereas the lower portion consists of less depleted mantle material (lherzolite). On the basis of geochemical and geophysical, especially seismic, observations, it is generally accepted that the typical oceanic upper mantle contains only a small amount of H_2O in the form of lattice impurities in minerals and contains little H_2O in the way of hydrated silicate phases except in the upper portion of the oceanic mantle where the infiltration of seawater through deep-cutting faults may hydrate the mantle locally. The principal hydrous minerals in H_2O-saturated peridotite at the pressure and temperature conditions of the oceanic upper mantle include serpentine, talc,

chlorite, and amphibole. Serpentine and chlorite contain 12.3 and 13 wt% H_2O, respectively, and therefore are important to the fluid budget in subduction zones (Pawley and Holloway, 1993; Ulmer and Trommsdorff, 1995). Antigorite that survives to ~200 km depth is replaced by another hydrous phase, phase A, which contains 11.8 wt% H_2O and is stable at higher temperature and pressure conditions than antigorite (Schmidt and Poli, 1998).

As the slab subducts, pressure and temperature increase, resulting in dehydration and metamorphic reactions of minerals in the slab. The stability of minerals has been determined through laboratory experiments and also theoretical treatment of thermodynamic data (e.g., Evans, 1977, 2004; Hacker et al., 2003a; Pawley and Wood, 1995; Schmidt and Poli, 1998; Ulmer and Trommsdorff, 1995). Recently, thermodynamic calculation codes for constructing phase diagrams for a given composition are readily available, including Perple_X (Connolly, 2005, 2009; Connolly and Petrini, 2002) and Thermo-Calc (Powell and Holland, 1988; Powell et al., 1998) (**Figure 3**). With these experimental results and thermodynamic calculations and numerical thermal modeling results, the stability of minerals in the subducting slab and the surrounding mantle is calculated (e.g., Hacker, 2008; Hacker et al., 2003b; van Keken et al., 2011; Wada et al., 2012) (see **Section 7.09.4.5**). Using Perple_X, the values of thermodynamic properties, such as density, thermal expansivity, and heat capacity, are also calculated at a given pressure and temperature and a given composition (e.g., Nakagawa et al., 2009). Further, by using numerical thermal models and experimentally determined solidus (e.g., Green, 1973; Grove et al., 2006; Kushiro et al., 1968; Schmidt and Poli, 1998) or thermodynamic calculation codes, such as MELTS (Asimow et al., 2004; Asimow and Ghiorso, 1998; Ghiorso and Sack, 1995; Smith and Asimow, 2005), the location and the degree of melting in subduction zones are also calculated (e.g., Grove et al., 2006, 2009; Hebert et al., 2009). Model-predicted thermopetrologic structure of subducting slabs is discussed in Sections 3.4 and **7.09.4.6**.

The lithologic and metamorphic structures of subducting slabs at shallow and intermediate depths have been inferred from seismic-wave dispersion and scattering. Dispersion of seismic waves occurs due to the frequency dependence of seismic waves and is particularly sensitive to the presence and physical state of subducted oceanic crust at high frequencies (0.5–10 Hz) (Abers, 2000). Scattering occurs due to three-dimensional heterogeneities in the medium that the waves travel. Where resolution allows, dispersion measurements and scattering wave fields image a low-velocity layer (LVL) along the top of the subducting slab, which is commonly interpreted to be the basaltic or metastable gabbroic oceanic crust that has not transformed to eclogite (Fukao et al., 1983) (**Figure 4**). As summarized by Wada and Wang (2009), an LVL is observed to 45–75 km depths in Cascadia (Bostock et al., 2002; Cassidy and Ellis, 1993; Rondenay et al., 2008) (**Figure 4(b)**) and to ~60 km depth in Nankai (Hori, 1990; Hori et al., 1985; Ohkura, 2000), both of which are warm-slab subduction zones. In other subduction zones, the downdip extent of an LVL varies from 120 km depth in northern Chile (Bock et al., 2000; Yuan et al., 2000) and Alaska (Abers et al., 2006; Rondenay et al., 2008) (**Figure 4(a)**) and 100–150 km depths in NE Japan (Kawakatsu and Watada, 2007; Matsuzawa et al., 1986). However, in NE Japan, seismic tomography by Tsuji et al. (2008) images an LVL disappearing at depths of 70–90 km, which is interpreted to indicate the dehydration of the subducted crust. With this interpretation, the deeper extension of the LVL represents the subducted crust with a less water content, which has not transformed into eclogite (Nakajima et al., 2009; Tsuji et al., 2008). A body wave dispersion study by Abers (2005) indicates an LVL extending to at least 150 km depth in Nicaragua, the Aleutians, Alaska, the Marianas, Kamchatka, and NE Japan, although the exact maximum depth is undetermined because of the large depth bins used in the study. Abers (2005) showed a substantial variation in waveguide velocities from arc to arc, correlating with slab dip but not with other subduction-related parameters (e.g., plate age, convergence velocity, and thermal parameter). Abers (2005)

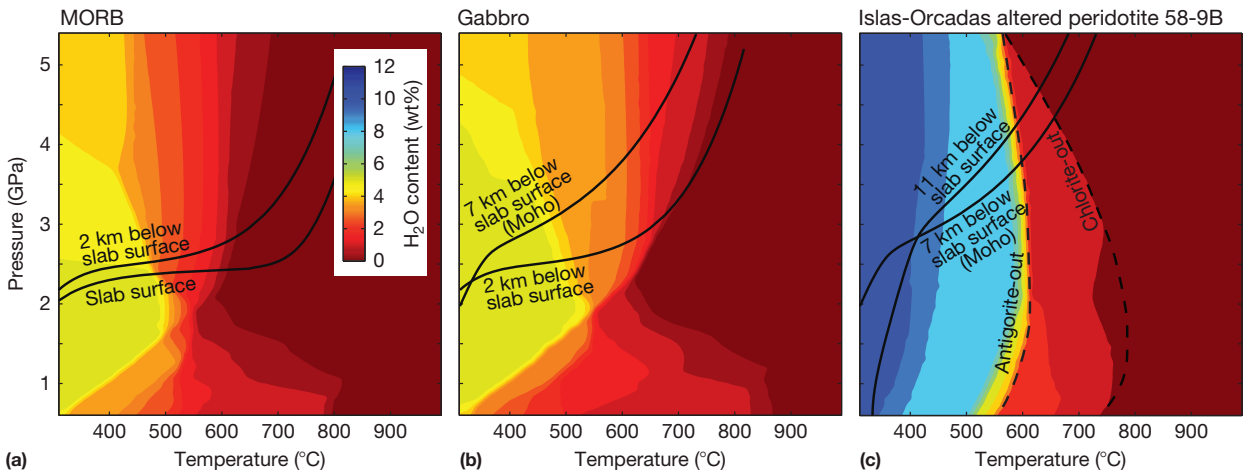


Figure 3 Contents of mineralogically bound H_2O at saturation calculated with Perple_X for (a) MORB, (b) gabbro, and (c) peridotite compositions. Reproduced from Wada I, Behn MD, and Shaw AM (2012) Effects of heterogeneous hydration in the incoming plate, slab rehydration, and mantle wedge hydration on slab-derived H_2O flux in subduction zones. *Earth and Planetary Science Letters* 353: 60–71.

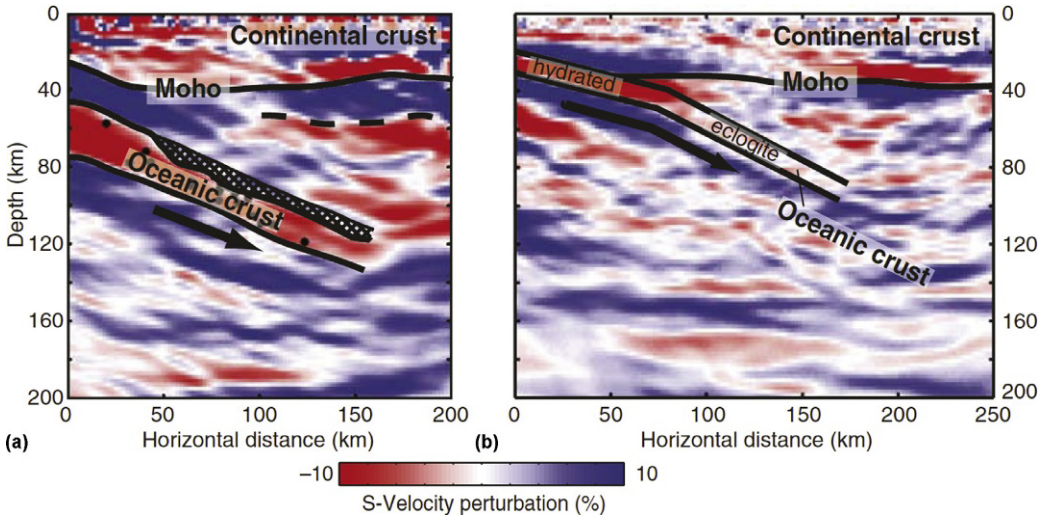


Figure 4 Scattered wave field across (a) central Alaska and (b) Cascadia overlain by interpretations (Rondeney et al., 2008). Red and blue regions indicate slower and faster shear-wave perturbation, respectively, relative to a 1-D background model. Reproduced from Rondenay S, Abers GA, and Van Keken PE (2008) Seismic imaging of subduction zone metamorphism. *Geology* 36: 275–278.

favored fluids or hydrous minerals as opposed to variable metastability of anhydrous gabbro as an explanation pointing out that the variability with dip is consistent with fluids from shallow slabs being driven into the mantle wedge while for steeper slabs, the fluid preferentially migrates up the sediment layer. Although the downdip extent of an LVL at a given subduction zone varies depending on the type of observations, there is a trend in which the LVL extends deeper in the older and colder slabs, such as in northern Chile, Alaska, and NE Japan, than in younger and warmer slabs, such as in Cascadia and Nankai. As discussed by Wada and Wang (2009), this trend agrees with petrologic predictions for the subducting slabs based on thermal models (e.g., Hacker et al., 2003b; Peacock and Wang, 1999; Wada and Wang, 2009) (**Section 7.09.4.5**). Interestingly, the studies by Kawakatsu and Watada (2007) and Tsuji et al. (2008) reveal another LVL above the top of the slab at depths of 100–120 and 80–130 km, respectively. This LVL is interpreted as a hydrated mantle layer at the base of the mantle wedge.

In Nicaragua, relatively low seismic velocities (e.g., Ranero et al., 2003) and high electrical resistivity (e.g., Key et al., 2012) are observed in the upper portion of the upper oceanic mantle prior to subduction, indicating hydration. Deep-cutting faults that result from tectonic extension near the spreading center and/or bending at the outer rise along convergent margins are proposed as the fluid pathways to the upper mantle although how fluids migrate downward to the upper mantle depth is debated. A recent thermomechanical modeling study (Faccenda et al., 2012) indicates that the oceanic mantle may become hydrated after subduction due to the upward migration of fluids along a zone of neutral pressure gradients that develop in the subducting mantle during slab bending (see **Section 7.09.5.6** for more discussions on plate bending). A seismic tomography study (Syracuse et al., 2008) shows anomalously low velocities in the subducting slab, extending to depths of 140 km beneath Nicaragua. Beneath Hokkaido in NE Japan, a narrow zone of low velocities is also observed in the subducting mantle (Nakajima et al., 2009). At both subduction zones, the low-velocity regions in the subducting mantle are interpreted to indicate the hydrated (serpentinized) oceanic mantle. Beneath Hokkaido, the low-velocity zone coincides with the lower plane of the double seismic zone, supportive of dehydration embrittlement as the intraslab earthquake mechanism. Serpentine and chlorite in the serpentinized oceanic mantle contain 12–13 wt% H_2O (see **Section 7.09.2.3**) and are stable to greater depths than most hydrous phases in the subducting crust. These observations help understand the distributions of hydrous phases and free aqueous fluids in the subducting slab, volatile cycle and budget in subduction zones, and their implications for subduction zone processes such as intraslab earthquakes. The metamorphic state of the subducting slab depends strongly on the degree and extent of hydration in the incoming oceanic plate, which is discussed in **Section 7.09.4.6** along with the thermopetrologic modeling of subducting slabs using coupled kinematic–dynamic models.

7.09.2.3 Slab Dehydration and Melting: Geochemistry of Arc Lava and the Location of the Arc

The addition of aqueous fluids released by the dehydrating subducting material into the hot region of the overlying mantle wedge triggers hydrous melting by lowering mantle solidus (Asimow et al., 2004; Gaetani and Grove, 1998; Hirth and Kohlstedt, 1996; Leeman, 1996; Pawley and Holloway, 1993; Ulmer, 2001). The role of water in melting at arcs has been studied extensively, dating back to Green and Ringwood (1968), and there are several reviews covering this topic (e.g., Eiler, 2003; Elliott, 2003; Hirschmann, 2006; Poli and Schmidt, 2002). The mechanism of hydration melting is supported by the relatively high water content of up to ~6 wt% H_2O (e.g., Cooper et al., 2012; Johnson and Plank, 1999) and enrichment of fluid-mobile elements, such as boron (B), in arc lavas (Ishikawa and Nakamura, 1994; Leeman, 1996). The rate

of magma production is thus expected to depend strongly on the amount of fluid fluxing from the subducting slab. Variation in the enrichment of B in arc lavas examined at several margins indicates greater fluid flux from the subducting slab (and sediments) for subduction zones with older slabs (Leeman, 1996). There is also some evidence for (adiabatic) decompression melting of the upwelling mantle in the volcanic arc of some subduction zones, such as the southern Marianas (e.g., Kohut et al., 2006).

The emerging view is that the geochemistry points to both volatile-induced melting in the mantle wedge and melting of subducting sediments and oceanic crust in subduction zones. There are extensive reviews by Elliott (2003) and Gaetani and Grove (2003) that cover work that is beyond the scope of this chapter. In some subduction zones, such as Cascadia, Nankai, the Aleutians, and Ecuador, in addition to flux melting of the mantle, partial melting of the top part of the subducting crust is inferred from a low yttrium (Y) and ytterbium (Yb) content and high strontium/yttrium (Sr/Y) ratio (e.g., Bourdon et al., 2002; Defant and Drummond, 1993; Morris, 1995; Yogodzinski et al., 1995). Rocks that form from this type of magma are referred to as adakites. Melting of the subducting crust requires very high temperatures and is thought to occur when the subducting slab is young (<25 Ma) and warm and/or the rate of shear heating is high (Peacock et al., 1994). The elevated level of trace elements such as beryllium (Be) and thorium (Th) in some arc lavas also suggests partial melting of subducted sediments (Johnson and Plank, 1999, references therein). The presence of ^{10}Be, which is produced only in the upper atmosphere with a half-life of 1.6 Myr, is a strong evidence that some oceanic sediments are recycled through subduction zones, travel through the mantle wedge, and wind up in the source region of arc lavas (Morris et al., 1990). Furthermore, the short half-life places a tight constraint on the transport of material through the subduction/arc system. Slab surface temperature inferred from the H_2O/Ce ratio of the melt inclusions of arc lavas globally ranges from 730 to 950 °C, which can cause melting of H_2O-saturated basaltic crust (Cooper et al., 2012; Plank et al., 2009; Skora and Blundy, 2010).

One of the intriguing observations related to arc volcanism is the location of the arc relative to the subducting slab. England et al. (2004) determined the depth to the top of intraslab seismicity beneath the arc for 24 margin segments globally using teleseismically determined intraslab earthquake hypocenter locations. The depth ranges from 65 to 130 km with the mean of ~100 km. About 54% of them fall in the range of 90–100 km, and 88% of them are in the 80–120 km range. There is a tendency of the slab to be at these depth ranges. Syracuse and Abers (2006) adjusted the subarc slab depth to account for errors in hypocenter locations, using the same earthquake catalog as in England et al. (2004) and delineating the slab surface based on the distribution of intraslab earthquakes. In general, for the 24 margin segments investigated by England et al. (2004), the subarc slab depths determined by Syracuse and Abers (2006) are systematically deeper by up to ~40 km. The subarc slab depths have also been determined independently using local seismic networks in a number of subduction zones, which indicate local variations in the depths, but overall the volcanic arc to form where the slab is about 100 km deep. Through the comparison of the subarc slab depth with a range of subduction parameters, both England et al. (2004) and Syracuse and Abers (2006) suggested that the kinematics of the subducting slab is at least partly responsible for the small variations in the subarc slab depth. The migration path of aqueous fluids and melts is likely to affect where the arc volcanoes form. However, what regulates the location of arc relative to the subducting slab and the configuration of the subduction zone in general is unclear at present.

7.09.2.4 Subduction Thrust Earthquakes, Episodic Tremor and Slip, and Intraslab Earthquakes

In subduction zones, earthquakes occur along the plate interface (interplate or subduction thrust earthquakes), in the overriding plate (mostly in the crust and thus often referred to as crustal earthquakes), and in the subducting slab (intraslab earthquakes). Subduction thrust earthquakes often involve large rupture zones and cause devastating damage and loss of life (e.g., the 2004 M_w 9.1 Sumatra earthquake and 2011 M_w 9.0 Tohoku-Oki earthquake). Based on fault mechanics (e.g., Scholz, 1990), the spatial limit of subduction thrust earthquake rupture zones corresponds to transition from stable sliding (aseismic) to stick-slip (seismic) frictional behaviors. One of the outstanding questions of what controls the change in the sliding behavior of the interface is beyond the scope of this paper, but it is important to point out that the thermal structure of the subducting material along the plate interface is an important factor that controls the sliding behavior (e.g., Ruff and Tichelaar, 1996) (see also Volume 4 of this treatise for earthquake seismology).

In northern Cascadia, nonvolcanic low-frequency seismic tremor occurs episodically around the tip of the mantle wedge corner in the forearc region (Kao et al., 2005; La Rocca et al., 2009; McCausland et al., 2005; Rogers and Dragert, 2003). A similar tremor activity is also detected near the mantle wedge tip in Nankai (Obara, 2002; Shelly et al., 2006). In both Cascadia and Nankai, slow slip events occur along the plate interface downdip of the locked seismogenic portion of the subduction interface concurrently with the tremor (Dragert et al., 2001; Hirose and Obara, 2005, 2006; Ozawa et al., 2002; Wang et al., 2008). The tremor and slow slip are together referred to as episodic tremor and slip (ETS). Detailed studies of the waveforms of tremor in Cascadia (Wech and Creager, 2007) and Nankai (Shelly et al., 2006, 2007) indicate shear slip, but it is not known whether both tremor and slow slip are caused by interface slip only. In both Cascadia and Nankai, seismic tomography indicates the presence of fluid in the ETS region, and thus, ETS may be associated with the shallow dehydration of the warm subducting slabs (Kao et al., 2005; Shelly et al., 2006). Rate- and state-dependent friction laws have also been applied to explain the initiation and propagation of slow slip (e.g., Liu and Rice, 2005, 2007). ETS-like events are also reported to occur in Mexico (Larson et al., 2007; Payero et al., 2008), Alaska (Ohta et al., 2006), and Costa Rica (Larson et al., 2007). See **Chapter 4.17** for further studies on slow slip.

Intraslab earthquakes occur over a wide depth range, extending to as deep as the base (660 km depth) of the mantle

transition zone in some subduction zones (Frohlich, 1989, references therein). The depth range of intraslab earthquakes varies greatly among subduction zones, but the maximum depth of intraslab seismicity tend to be greater for colder slab (Kirby et al., 1996; Wada and Wang, 2009).

Intraslab earthquakes at intermediate depths (~50–300 km depth) occur in two subparallel bands (double seismic zone) in many subduction zones (e.g., Brudzinski et al., 2007). They are thought to represent seismicity in the subducting crust and upper mantle separated by an aseismic core (Hacker et al., 2003b), which tends to be thicker for older and colder slabs (e.g., Brudzinski et al., 2007). Tectonic stresses associated with unbending of the slab near the trench are proposed as one possible cause of the double seismic zones (e.g., Igarashi et al., 2001; Wang, 2002). While tectonic stresses cause earthquakes, relatively high confining pressure at depths prevents seismic failure. Dehydration embrittlement of the subducting crust and mantle is proposed to facilitate intermediate-depth earthquakes (e.g., Kirby et al., 1996; Peacock, 2001; Yamasaki and Seno, 2003). This mechanism refers to the process by which fluid released during the dehydration of hydrous minerals in the subducting slab elevates pore fluid pressure, reducing the effective pressure and facilitating seismic rupture (e.g., Kirby et al., 1996). According to this hypothesis, the presence of fluid causes rock failure to occur seismically, controlling the distribution of intraslab earthquakes.

For deep earthquakes (300–660 km depths), the mineral phase transformations of metastable olivine into wadsleyite or ringwoodite arose as an intriguing possible mechanism, referred to as transformational faulting, following the Green and Kirby laboratory experiments (cf. Green and Burnley, 1989; Kirby et al., 1991, 1996; Stein and Rubie, 1999). During the phase transformation, microinclusions of newly transformed minerals preferentially develop in a direction perpendicular to the axis of maximum compression (Green and Burnley, 1989). The transformation is accompanied by volume reduction, generating tensile stress in the direction of microinclusion development, which promotes nucleation of faults (Green and Burnley, 1989; Kirby et al., 1991). Other specific mechanisms that have been proposed for deep earthquakes include dehydration or amorphization of hydrous phases (e.g., Meade and Jeanloz, 1991) and shear instability (Kanamori et al., 1998; Ogawa, 1987). Although the mechanisms of both intermediate-depth and deep intraslab earthquakes are unclear, their generation appears to be associated closely with the slab's thermal structure and metamorphic processes. For further studies on deep intraslab earthquakes, see **Chapter 4.13**.

It was recognized that earthquake focal mechanisms of the deepest earthquakes reaching to 410–660 km depths are aligned with the axis of compression along the dip of the plane delineated by the hypocenters (cf. Isacks and Molnar, 1971; see **Chapters 4.13** and **4.18**), and this was taken to be evidence that the downgoing lithosphere meets some form of resistance (e.g., Richter, 1973; Vassiliou and Hager, 1988). More recently, Alpert et al. (2010) used moment tensor solutions to infer the orientation of the compressive stress axes in the subducting slab globally (**Figure 5**). They found that at intermediate (100–350 km) depths, the slab is predominantly under extension in the east-directed subduction zones and compression in the west-directed subduction zones. Using moment tensor solutions and employing a global slab geometry model (slab contours of the regionalized upper mantle (RUM) seismic model of Gudmundsson and Sambridge, 1998), Bailey et al. (2012) performed systematic global analyses of the stress orientation relative to the slab dip. They found that the state of stress in the subducting slabs can be categorized into three groups as proposed by Isacks and Molnar (1971), intermediate-depth extension, intermediate-depth extension–deep compression, and intermediate-depth deep compression, with some exceptions that are likely related to slab bending, oblique subduction, or mechanical heterogeneity. These observations indicate variability, though systematic, in the amount of forces that slabs experience, particularly in the shear resistance of the surrounding mantle.

The average downdip strain rate in subducting slabs has been estimated based on the rate of accumulation of seismic moment in the Wadati–Benioff zones (Bevis, 1986; Fischer and Jordan, 1991; Holt, 1995), assuming that the amount of aseismic deformation is small and that the viscous deformation is parallel with the brittle layer with the same deformation rate tensor. Because the estimate does not include the amount of aseismic deformation, it represents the lower bound of the total strain rate in the slab. Bevis (1988) estimated that in the depth range of 75−175 km, the average strain rate in subducting slabs is $10^{-15}\,s^{-1}$. This compares with an asthenospheric strain rate of $8 \times 10^{-15}\,s^{-1}$, assuming that most of the strain is accommodated within the top 200 km of the asthenosphere (e.g., take a characteristic plate velocity of 5 cm per year divided by a thickness of 200 km), leading to the rather surprising suggestion that the difference in deformation rates between the asthenosphere and subducting slabs is no more than a factor of 8. Taking the average descent rate of slabs, this deformation rate of $10^{-15}\,s^{-1}$ corresponds to a total accumulated strain in the depth range of 75–175 km of 0.06. Because slabs are generally in downdip compression or downdip extension, the sum of the seismic moments provides an estimate for the average change in the length of a slab caused by seismicity (Bevis, 1988). Fischer and Jordan (1991) found that seismic data require a thickening factor of 1.5 or more in the seismogenic core of the slab in central Tonga. Holt (1995) argued that the average seismic strain rate in Tonga represents as much as 60% of the total relative vertical motion between the surface and 670 km. These studies indicate that subducting slabs undergo a significant amount of deformation.

Creager et al. (1995) showed that the concave oceanward bend in the Bolivian orocline (South American slab) forces along-strike compression in the subducting slab. Their calculations of membrane strain rates in an assumed continuous slab demonstrate that this can be accommodated by either 10% along-strike compressive strain or geometric buckling. They present evidence for complex deformation in the deep slab that could contribute to the conditions for very large earthquakes by a dramatic increase in deformation rate and anomalous advective thickening of a proposed overdriven olivine wedge. This provides a mechanism for local thickening of the olivine wedge and allows transformation faulting to occur over a much larger volume than for normal subduction. Based on the deformation rate in the subducting slab inferred from seismic moment, the slab appears to undergo a significant

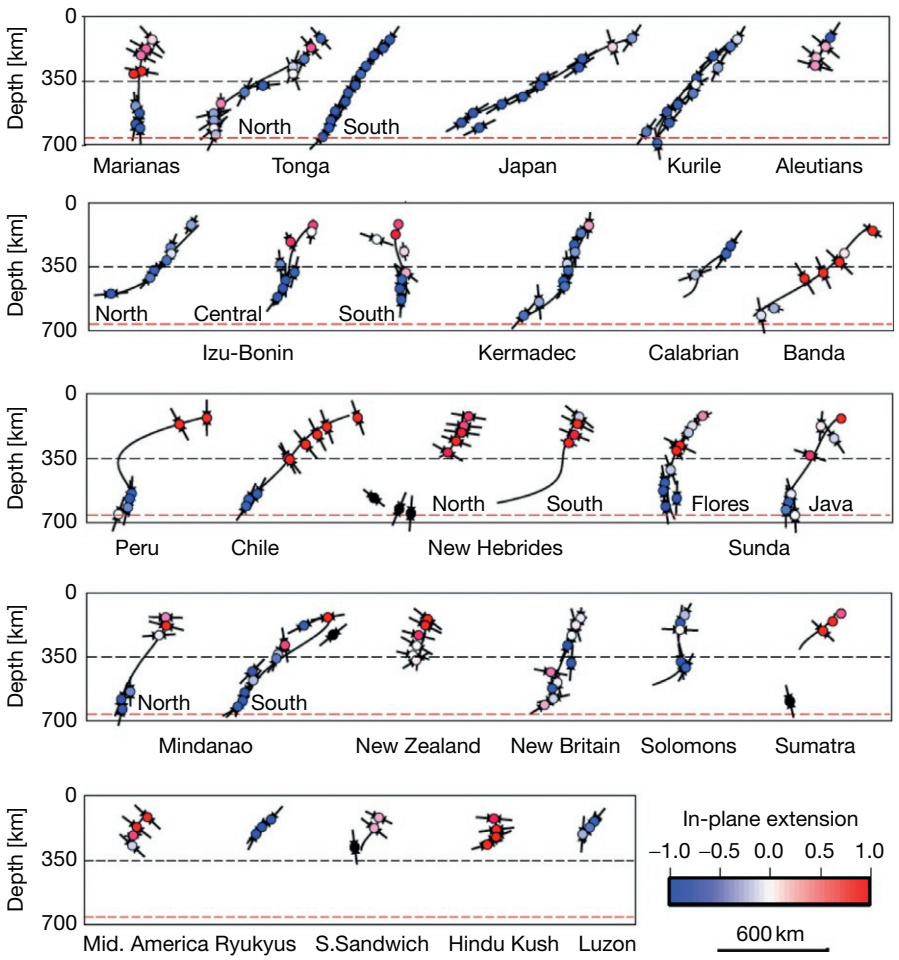


Figure 5 A global summary of variations in in-plane extension (color of the circles) calculated from moment tensor solutions (Alpert et al., 2010). Arrows around the circles indicate the orientation of the compressional axes. Black circles represent solutions that cannot be used to infer the state of stress. The line represents the slab geometry delineated from the distribution of intraslab seismicity. Reproduced from Alpert LA, Becker TW, and Bailey IW (2010) Global slab deformation and centroid moment tensor constraints on viscosity. *Geochemistry, Geophysics, Geosystems* 11: Q12006.

amount of deformation. It would be easy to dismiss any one of these observations taken by itself; however, the number and diversity of the observations that slabs are not significantly stronger than ambient mantle suggest that this should receive serious consideration. A weak slab is not necessarily inconsistent with the laboratory observations. While the viscosity of olivine is a strong function of temperature, it is also a strong function of grain size (Riedel and Karato, 1997). Grain-size reduction resulting from recrystallization of minerals at phase boundaries may counterbalance the effects of temperature on viscosity; however, this is not an obvious solution because grain reduction occurs in dislocation creep, whereas grain weakening occurs in diffusion creep. One of the proposed mechanisms that overcome this paradox is the grain-damage and pinning mechanism proposed by Bercovici and Ricard (2012), which results in both grain-size reduction and grain weakening during diffusion creep and thus shear localization. In addition, the brittle component of the slab may be mechanically weakened by fractures and faults, and thus, the single-crystal measurements of viscosity may not correctly describe the deformation of the slab.

As one example where deep mantle flow is almost certainly impacting the deformation of a slab, Gurnis et al. (2000a) showed that the Tonga slab is being deformed by a large-scale upwelling associated with the Pacific superplume. There is no substantial aseismic penetration into the lower mantle beneath Tonga, consistent with initiation of subduction during the Eocene. Both the pattern and amount of deformation and the seismic energy release in the Tonga slab show that it is deforming faster and has accumulated more deformation than any other slab. Gurnis et al. (2000a) argued that the strong deformation of the Tonga–Kermadec slab in the transition zone is a result of the short subduction history (40 Myr) and the upward flow in the lower mantle from a broad upwelling associated with the Pacific superplume. The lack of deep aseismic extension of the Tonga slab into the lower mantle is consistent with the short subduction history and the deep mantle flow. While recognizing the importance of regional plate motions, especially the opening of the Lau Basin, Gurnis et al. (2000a) argued that the structure and geometry of the Tonga–Kermadec slab are best explained with controls from deep and shallow forces.

7.09.2.5 Shear-Wave Splitting in Subduction Zones

Seismic anisotropy in the upper mantle is generally assumed to be the result of lattice-preferred orientation (LPO) of mantle minerals, such as olivine, that develops during dislocation creep and causes orthogonal components of shear waves to travel at different velocities (shear-wave splitting) (see **Chapter 2.20**). Shear-wave splitting observations provide a fast polarization direction and a splitting time, reflecting the orientation, organization, and strength of the fabric (Savage, 1999; Silver, 1996). Conventionally, olivine a-axes (the seismically fastest axis) are assumed to be aligned with the shear direction (A-type olivine fabric) (e.g., Karato and Wu, 1993), leading to the interpretation that the fast direction is aligned with the direction of mantle flow. However, deformation experiments show that the LPO is sensitive to the presence of water, temperature, and stress. For example, in the presence of H_2O as lattice impurities under relatively high stress and low temperature, the LPO of olivine results in a fast direction that is normal to the convergence direction (B-type olivine fabric) (e.g., Jung and Karato, 2001; Katayama and Karato, 2006). Jung and Karato (2001) suggested that the B-type olivine fabric dominates in the mantle wedge corner close to the trench where water-rich conditions should prevail, leading to the trench-parallel fast direction in the wedge with trench-normal mantle flow. The sensitivity of olivine fabrics to water, temperature, and stress makes the interpretation of the seismic-wave splitting observations challenging. Other factors, such as the presence of faults, fracture, and lenses of melt (e.g., Healy et al., 2009; Spiegelman, 2003), and the formation of hydrous phases (e.g., Jung, 2011; Jung and Karato, 2001; Katayama et al., 2009; Lassak et al., 2006) can also influence seismic anisotropy, and the interpretation of the observations is not straightforward. Karato et al. (2008) provided a comprehensive review on the seismic anisotropy of the upper mantle, and Long (2013) provided a summary of a range of constraints that are currently available on subduction zone anisotropy.

Shear-wave splitting orientations show a variety of fast directions among different subduction zones and within a single arc (**Figure 6**). In the Tonga subduction zone (Bowman and Ando, 1987; Fischer et al., 2000; Smith et al., 2001), shear-wave splitting results vary from trench-parallel fast directions near the trench to convergence-parallel fast directions closer to the back-arc (Long and Silver, 2009; Long and Wirth, 2013). In NE Japan, shear-wave splitting observations also range from trench-parallel near the trench to convergence-parallel in the back-arc (Hiramatsu and Ando, 1996; Nakajima and Hasegawa, 2004; Wirth and Long, 2010). The Japan subduction system is complex with two triple junctions and multiple subducting slabs. In addition, crustal anisotropy may be responsible for some local variability of splitting variations (e.g., Fouch and Fischer, 1996); however, there is agreement that mantle anisotropy is required to explain the majority of the shear-wave splitting results across the region (e.g., Fouch and Fischer, 1996; Long and van der Hilst, 2005; Nakajima and Hasegawa, 2004). Near Hokkaido, Nakajima and Hasegawa (2004) observed splitting variations that are consistent with north–south shear in the overriding plate or the occurrence of B-type olivine fabric in the mantle wedge. In Ryukyu (Fouch and Fischer, 1996; Long and van der Hilst, 2005) and Kamchatka (Levin et al., 2004; Peyton et al., 2001), the fast direction rotates from trench-parallel near the trench to trench-orthogonal into the back-arc. Levin et al. (2004) interpreted these measurements as the result of a complex flow regime with mantle flow moving around the northern bending edge of the downgoing slab due to slab loss/detachment, possibly combined with deformation of material

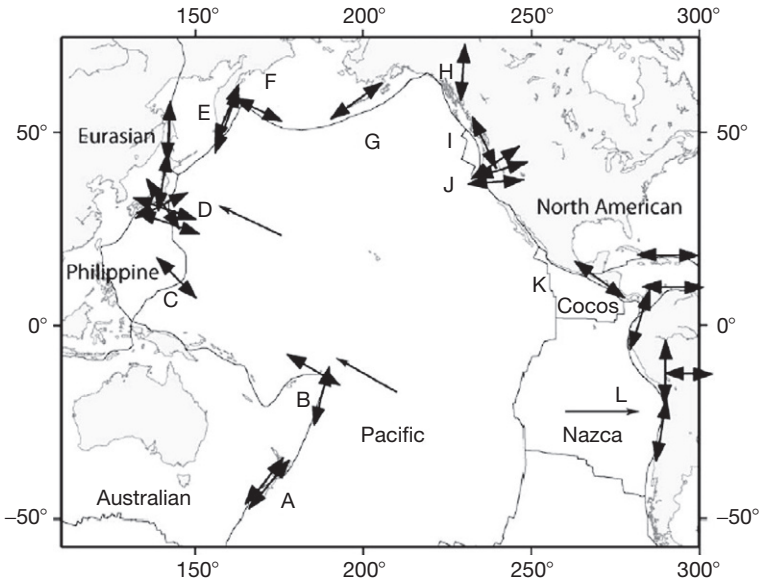


Figure 6 Summary of shear-wave splitting observations around the Pacific Rim illustrating the variety of observed fast directions near subduction zones from Lassak et al. (2006). Double-headed vectors indicate the average regional orientation of the fast polarization direction. Many regions exhibit fast direction populations with orientations both parallel and orthogonal to local trench strike. Reproduced from Lassak TM, Fouch MJ, Hall CE, and Kaminski E (2006) Seismic characterization of mantle flow in subduction systems: Can we resolve a hydrated mantle wedge? *Earth and Planetary Science Letters* 243: 632–649.

in the wedge. Lassak et al. (2006) suggested that B-type olivine fabric in the wedge may also be a possible explanation. South America also exhibits a range of shear-wave splitting directions (Anderson et al., 2004; Polet et al., 2000; Russo and Silver, 1994). Long and Silver (2009) summarized the shear-wave splitting observations presented in the earlier-mentioned references and suggested two reasons why anisotropy in the mantle wedge is difficult to interpret. First, the observed splitting patterns in the wedge are variable with some subduction zones exhibiting a transition from trench-parallel anisotropy close to the trench to trench-perpendicular anisotropy farther away from the trench (e.g., Ryukyu, the Marianas, and Tonga), but Kamchatka exhibits the opposite pattern. Secondly, the conditions in the mantle wedge are such that it is likely that olivine could be in A-type LPO in some places and B-type LPO in other places. This would produce a 90° rotation in the observed shear-wave splitting orientation even if the strain rate field remained uniform. A number of studies have used mantle flow models to predict the LPO development and resulting shear-wave splitting (e.g., Fischer et al., 2000; Fouch et al., 2000; Hall et al., 2000; Kaminski and Ribe, 2001; Kneller et al., 2008; Lassak et al., 2006; Long and Silver, 2008). These studies use either local flow directions or finite strain directions as a guide for the orientation of crystallographic axes. For example, Kaminski and Ribe (2001) had developed a model for the evolution of LPO in olivine aggregates that accounts for intercrystalline slip and dynamic recrystallization. These shear-wave splitting observations and modeling results appear to point to a flow environment in subduction zones that is more complex than the simple 2-D corner flow. The presence of a relatively thin anisotropic layer within the mantle wedge in NE Japan has been inferred from anisotropic receiver function analysis (Wirth and Long, 2012), and the source of anisotropy in the mantle wedge may be highly localized.

There are observations that are difficult to explain by the B-type olivine fabric in the forearc mantle wedge. First, B-type fabric requires significant shear deformation in the stagnant mantle wedge. The numerical model of Wada et al. (2008) predicts little shear deformation in the stagnant wedge above the decoupled interface. Second, in the arc and back-arc regions of some subduction zones, such as Tonga (Smith et al., 2001), Ryukyu (Long and van der Hilst, 2006), and the Marianas (Pozgay et al., 2007), the fast direction is arc-parallel beneath the arc and the back-arc, where temperatures are too high and stresses are too low to allow the formation of B-type fabric. Third, in Cascadia and Nankai, where seismic observations indicate serpentinization/hydration of the mantle wedge, the convergence-parallel fast direction is observed (Currie et al., 2004a; Tono et al., 2009).

Contrary to the previously mentioned interpretations resorting to the anisotropy in the mantle wedge, Faccenda et al. (2008) argued that steeply dipping faults in the subducting slab and anisotropic hydrous minerals that form along the faults can explain the observed fast splitting directions in the forearc regions of most subduction zones. They also explained that the hydrated upper portion of the young Juan de Fuca slab in northern Cascadia is likely to be too thin to cause significant anisotropy and the convergence-parallel fast direction (Currie et al., 2004a) is likely to be caused by the asthenospheric mantle flow. This interpretation is also consistent with the convergence-parallel fast direction reported for Nankai (Tono et al., 2009) where the slab is also young. Anisotropic receiver function analyses and anisotropic P-wave tomography are used to obtain more direct constraints on the anisotropic structures of the subducting slab (Long, 2013), indicating the anisotropy of the subducting lithospheric mantle correlates with the fossil spreading direction in several subduction zones such as Central America and SW Japan (e.g., Song and Kim, 2012; Wang and Zhao, 2012).

7.09.2.6 Correlations among Slab Buoyancy, Kinematics, and Geometry

For the analyses of the correlations among various subduction parameters, such as convergence rate, slab age, and slab dip, we begin by focusing on the GPE associated with subducted slabs, which is the primary energy source driving mantle convection (e.g., Davies and Richards, 1992; Forsyth and Uyeda, 1975; Richter, 1977; Turcotte and Oxburgh, 1967). First, we estimate the thermal structure (hence buoyancy) that is coming into the trench as a result of the cooling of the oceanic plate. The thickness of an oceanic plate, h, is proportional to the square root of the time the plate spends at the surface. The constant of proportionality is the square root of the thermal diffusivity, κ. This time can be estimated by the length of the plate, L_p, divided by the plate velocity v_p. Hence,

$$h = \sqrt{\frac{\kappa L_p}{v_p}} \qquad [1]$$

(cf. **Chapter 7.04**). As the plate subducts, the cold material is denser than the warmer mantle surrounding it. Thus, the density difference between the slab and mantle will be $g\rho\alpha\Delta T$ where α is the coefficient of thermal expansion, ρ is the density, g is the gravitational acceleration, and ΔT is the average temperature difference between the interior mantle and surface. This buoyancy must be balanced by resistance of the mantle surrounding the slab, which has the form ηD_{ij}, where η is the coefficient of dynamic viscosity and D_{ij} is the deformation rate tensor (see **Chapter 7.02**). A crude estimate of the deformation rate is the velocity gradient, which we will assume goes from v_p at the plate surface to zero at some point near the middle of the mantle. Assuming L_p is approximately the same as the depth of the mantle and balancing the buoyancy force and viscous resistance, one obtains

$$v_p = L_p \left| \frac{g\rho\alpha\bar{\Delta}T\kappa^{1/2}}{2\eta} \right|^{2/3} \qquad [2]$$

With a reasonable choice of values for these parameters, one obtains a value for v_p on the order of 100 mm year^{-1}. This value is remarkably close to the velocity of oceanic plates with subduction zones (cf. Forsyth and Uyeda, 1975), especially when one considers the simplicity of the analysis.

For each of the major plates, Forsyth and Uyeda (1975) compiled plate velocity, area of plate, length of ridge, and trench and transform fault along the circumference of the plate boundary and calculated the relative importance of each contribution, concluding that the subducting slabs dominate the force balance. Taking their data (from **Table 1** of Forsyth

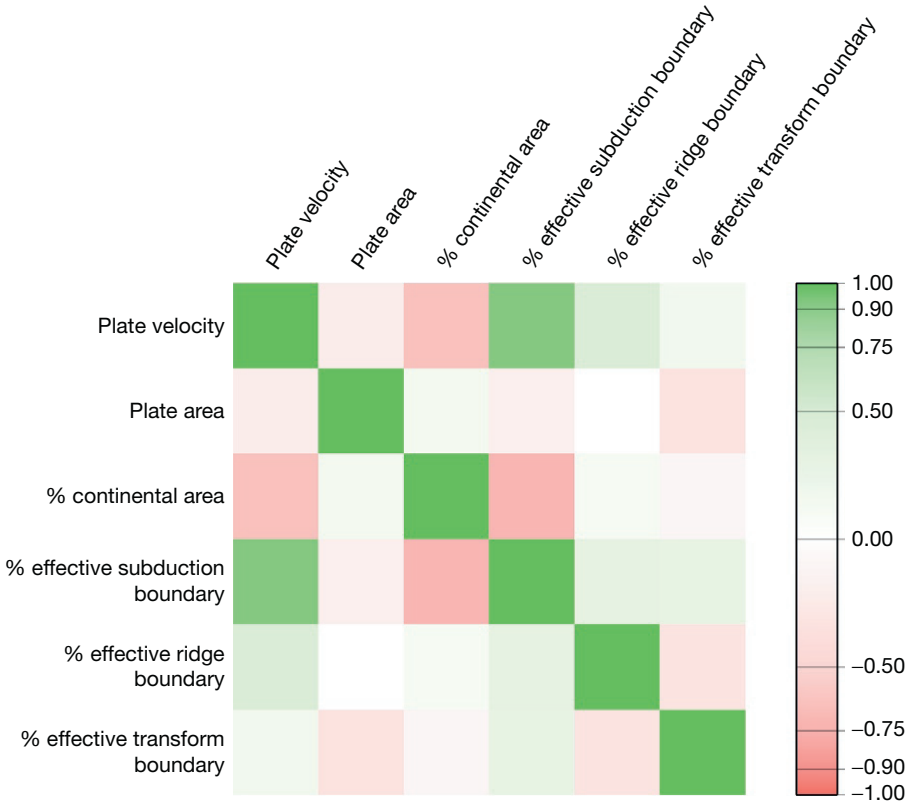


Figure 7 Color shading of the covariance matrix for selected plate geometry characteristics from Forsyth and Uyeda (1975) (see **Table 1** of Forsyth and Uyeda). The percentages of subduction zone, ridge, and transform fault boundary are the effective percentages (as defined by Forsyth and Uyeda), where the effects of boundaries that occur on opposing sides of the plate are assumed to cancel out and are not included in the percentage of length of that type along the boundary. Green is strong positive correlation; red is strong negative correlation.

and Uyeda, 1975), we plot the covariance matrix (**Figure 7**), which shows the strongest correlation of plate velocity with the length of the effective subduction boundary with a correlation coefficient of 0.92. The continental area has a second highest correlation coefficient of 0.64, and the length of the effective ridge boundary has the third highest correlation coefficient of 0.46. The total area of the plate and the length of the effective transform boundary and other observations tabulated by Forsyth and Uyeda (1975) show little correlation with plate velocity. (This figure is a compact representation of the graphs shown in Figures 5–9 of Forsyth and Uyeda, 1975.)

Significant attention has been given to correlating slab dip with a variety of observables related to subduction zones. A seminal compilation of 26 parameters measured or estimated for 39 modern subduction zones was undertaken by Jarrard (1986). Using these observations, Jarrard performed a multistep correlation analysis, finding that three independent variables, the age of the oceanic plate at the trench, the convergence rate of the plates at the trench (based on global plate reconstructions, e.g., Chase, 1978; Minster and Jordan, 1978), and intermediate slab dip (the dip angle between the trench and the slab at 100 km depth), correlate significantly with the other observables. There is an excellent correlation between the slab length and the product of the convergence rate and the age of the downgoing slab, consistent with the predictions of some slab thermal models (Molnar et al., 1979). We summarize the subduction zone parameter correlations by plotting a covariance matrix with nine of the best constrained observations from Jarrard's compilation plus the slab thermal parameter (slab age times convergence velocity) in **Figure 8**. There is also a strong positive correlation between the age of the plate at the trench and the age calculated at the slab tip, which is expected if the age of the slab tip is a linear function of the age of the incoming plate. This would be the case if the slab was rigid and the fact that this correlation is one indicator that slab deformation is limited. The other striking correlation is that between the age of the arc and the intermediate slab dip (trench to 100 km depth) and the deep slab dip (the average of the slab dip between 100 and 400 km). Trench motion relative to the overriding plate was also examined as a subduction parameter by Jarrard (1986), and its correlation with the slab dip was very weak at best. Jarrard (1986) found as many advancing as retreating trenches while Garfunkel et al. (1986) found that trench retreat predominates.

Slab dynamics and back-arc deformation style were reexamined by Heuret and Lallemand (2005) using relocated hypocenters (Engdahl et al., 1998), updated plate ages (Muller et al., 1997), and an updated global plate velocity model (Gripp and Gordon, 2002). After excluding regions where continental lithosphere, volcanic arc crust, or oceanic plateaus are being subducted, they divide the oceanic subduction zones into 159 segments with a uniform interval of two degrees of trench

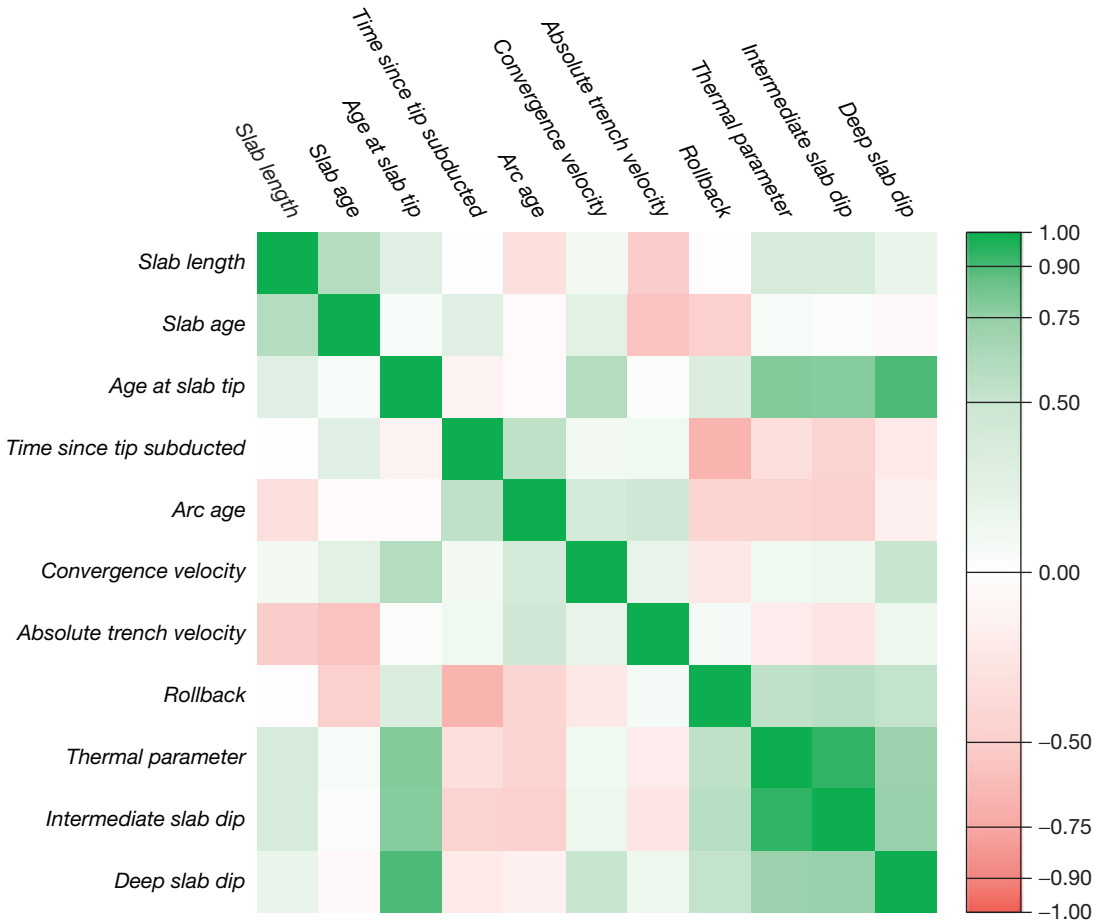


Figure 8 Covariance matrix for 9 observations compiled from 39 modern subduction zones (Jarrard, 1986). The color scale represents the value of the correlation coefficient, with +1.0 being strongly correlated and −1.0 being strongly anticorrelated. Rollback is the rate of trench migration estimated from the global plate models (Minster and Jordan, 1978). Velocity is the velocity of the plate at the trench from the same global plate models. Slab age is the age of the plate entering the trench, and slab tip age is calculated using the length of the slab and rate of change of the plate age for the plate entering the trench. The intermediate dip is calculated from the trench to the 100 km depth of the slab. The deep dip is calculated from the 100 to 400 km depth of the slab. The thermal parameter is the product of the slab age and the length of the slab.

length. Heuret and Lallemand (2005) found as many advancing as retreating trenches with slab motion occurring in two distinct populations: one with essentially zero trench velocity and one with a significant (-50 mm year^{-1}) component of retreat. As one might expect, they found back-arc extension when the trench is retreating and back-arc compression when the trench is advancing. The correlation between the absolute velocity of the overriding plate and the style of back-arc deformation from this compilation is consistent with Jarrard (1986). However, when correlating slab motion and age, Heuret and Lallemand's result (2005) is opposite to the result seen in many numerical models: older subducting plates are observed to be advancing, not retreating as is the case in many dynamic subduction calculations (e.g., Becker et al., 1999; Griffiths et al., 1995; Gurnis and Hager, 1988; Scholz and Campos, 1995). Heuret and Lallemand (2005) suggested that slabs are partially anchored because of the slabs' resistance to large-scale lateral flow. The degree of variability in the data suggests that local flow effects, such as small-scale flow in the vicinity of retreating slab edges or tears or more regional-scale flow associated with the shrinking of the Pacific Plate area (e.g., Garfunkel et al., 1986), may also significantly impact slab–trench dynamics. The statistic analyses by Lallemand et al. (2005) revealed that the slab dip correlates with the state of back-arc stress and the absolute motion of the overriding plate, but not with the magnitude of slab pull, the slab age, or the convergence rate.

Following a slightly different approach, Cruciani et al. (2005) reexamined the global correlation of slab dip with subduction zone observations using the RUM seismic model (Gudmundsson and Sambridge, 1998) to map out the slab geometry rather than using relocated earthquake hypocenters, as both Jarrard (1986) and Heuret and Lallemand (2005) did. Cruciani et al. (2005) constructed 164 trench-perpendicular cross sections through the RUM model using a digital seafloor age model for the age of the incoming slab (Muller et al., 1997) and a global plate velocity model for the convergence velocity (DeMets et al., 1994). There is more scatter in the Cruciani et al.'s (2005) compilation than the original compilation by Jarrard (1986), possibly because of the use of a seismic model as opposed to the more direct focal mechanism data, but more likely reflecting the actual complexity of slab geometries.

The product of slab age and plate convergence rate (thermal parameter) correlates with slab dip better than either slab age or plate convergence rate individually, although the correlation is weak (correlation coefficient of 0.45) and suggests that forces beyond the local subduction environment may be important. Hager and O'Connell (1981) showed that they could produce a reasonably good fit to slab dips simply with global plate motions and large-scale return mantle flow, without a detailed model of the deformation in the subduction zone, suggesting that global-scale mantle flow may play a role in slab geometry.

7.09.2.7 Geoid and Topography

The association between local maxima in the long-wavelength component of the Earth's gravitational potential field (geoid) and subduction zones has been recognized by many authors (Chase, 1979; Crough and Jurdy, 1980; Davies, 1981; Hager, 1984; Kaula, 1972; Mckenzie, 1977; Runcorn, 1967) (see **Chapter 3.02**). The geoid is the surface of constant potential energy (i.e., GPE plus centrifugal potential energy) that coincides (almost) with mean sea level over the oceans. (Mean sea level is not quite a surface of constant potential, due to dynamic processes within the ocean; however, for illustrative purposes, will suffice.) The geoid is actually the height of the potential surface referenced to an oblate spheroid resulting from a hydrostatic spinning Earth (Nakiboglu, 1982).

In a static mantle, with no dynamic flow, the geoid over a positive mass anomaly is positive (**Figure 9**, left panel). In an isoviscous fluid (**Figure 9**, middle panel), the gravitational potential (or geoid) over a positive mass anomaly, such as a subducting slab, is negative (Richards and Hager, 1984). This is because the flow driven by the mass anomaly (blue in **Figure 9**) deforms the surface (and core–mantle boundary) creating the negative mass anomalies (i.e., long-wavelength depressions). The sum of the contributions from the positive and negative mass anomalies in the isoviscous fluid is negative (and small) (cf. Ricard et al., 1984; Richards and Hager, 1984). With a layered viscosity increase with depth (**Figure 9**, right panel), the amount of surface deformation from the same positive mass anomaly is reduced (i.e., the mass anomaly is supported by the stiff layer below) and the resulting anomaly can change sign depending on the viscosity structure (cf. Ricard et al., 1984; Richards and Hager, 1984) (**Chapters 7.02** and **7.04**).

The power spectrum of the geoid is dominated by the longest wavelengths and the power decreases at approximately the fourth power of the spherical harmonic degree (Kaula, 1968). This observation has become known as Kaula's rule. As a result, the analysis of all but the longest components of the geoid requires spatial filtering. **Figures 10** and **11** compare the global geoid from the NIMA/NASA EGM96 geopotential model (Lemoine et al., 1997) that has been band-pass-filtered using two bands. The first band (**Figure 10**) cuts wavelengths greater than 15 000 km and smaller than 4000 km and passes wavelengths between 10 000 and 5000 km. The second band (**Figure 11**) cuts wavelengths greater than 5000 km and passes wavelengths smaller than 4000 km. **Figure 10** is comparable to the spherical harmonic degree 4–9 geoid that has been shown to correlate well with subduction zones (Hager, 1984), and the correlation between geoid highs and subduction zones is evident with the Aleutian Arc being a notable exception. Short-wavelength geoid anomalies, like those presented in **Figure 11**, have been presented in trench-perpendicular profiles (e.g., Billen et al., 2003; Davies, 1981; King and Hager, 1994; Zhong and Gurnis, 1992), and the geoid and topography at this wavelength have been used to constrain dynamic subduction zone models (Chen and King, 1998; King, 2002; King and Hager, 1994; Zhong and Gurnis, 1992, 1994, 1995) (see **Section 7.09.5.2**). Richards and Engebretson (1992) pointed out that the long-wavelength geoid correlates with the general pattern of fast seismic velocities in the Pacific.

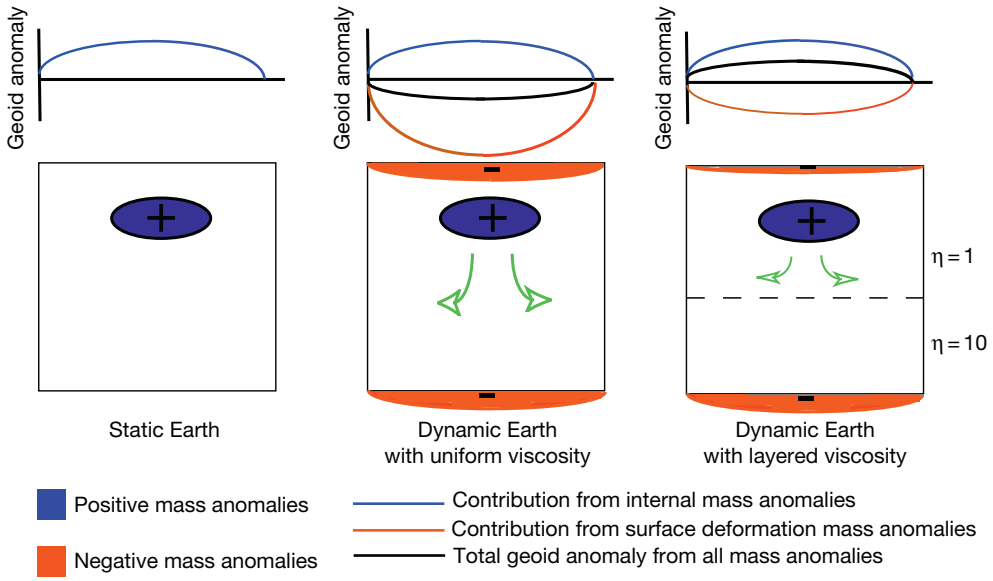


Figure 9 Schematic representation of geoid anomalies over a positive mass anomaly for a static (left), isoviscous (center), and layered viscosity (right) mantle. The blue line is the component of the anomaly due to the internal mass, the red line is the component of the anomaly due to the surface deformation, and the black line is the sum of the components.

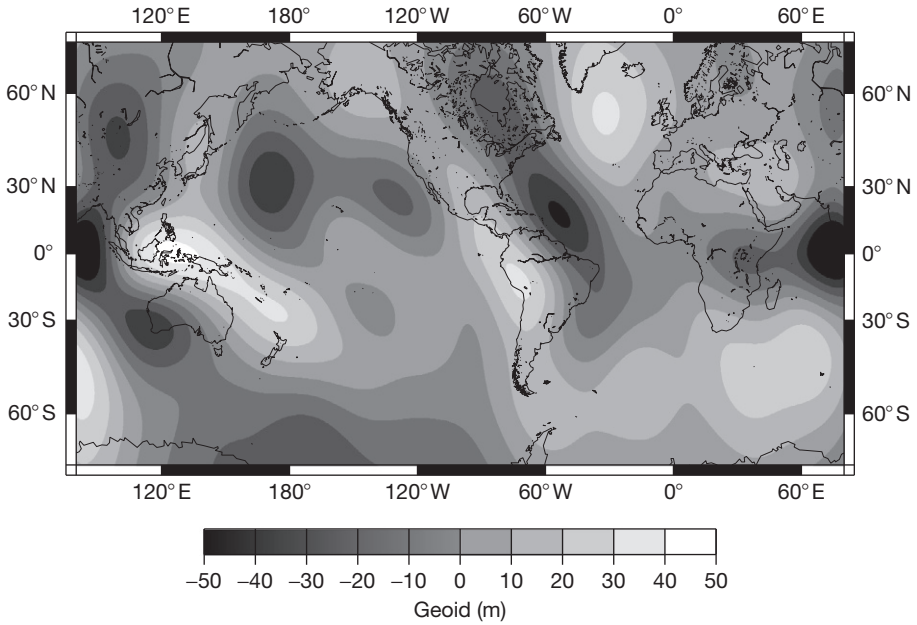


Figure 10 Global geoid from EGM96 (Lemoine et al., 1997) band-pass-filtered passing wavelengths smaller than 15 000 km and cutting wavelengths greater than 5000 km. These parameters are comparable to the degree 4–9 'slab' geoid used by Hager (1984). Routines in the GMT software package were used to filter the data (Wessel and Smith, 1991). Reproduced from King SD (2002) Geoid and topography over subduction zones: The effect of phase transformations. *Journal of Geophysical Research-Solid Earth* 107.

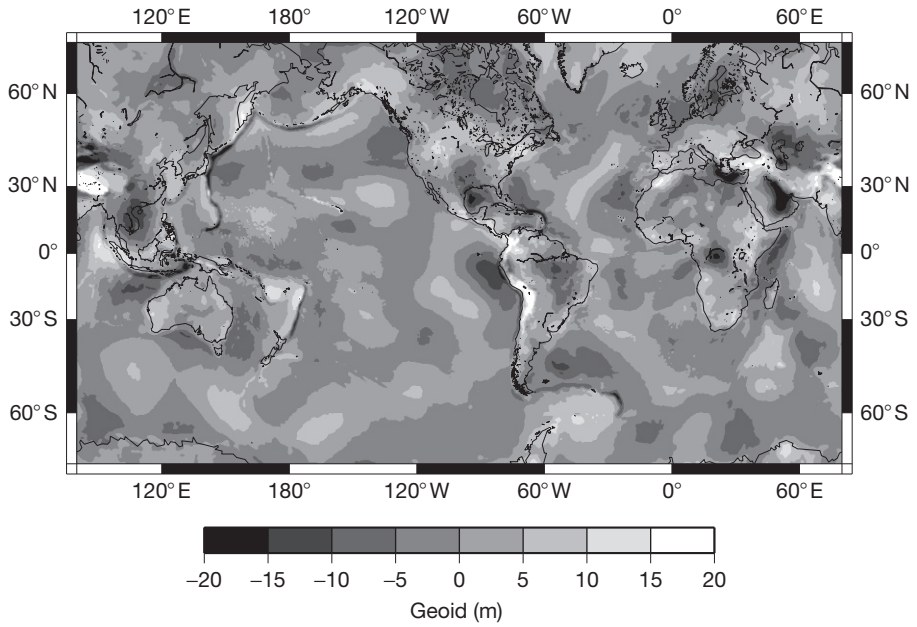


Figure 11 Global geoid from EGM96 (Lemoine et al., 1997) band-pass-filtered passing wavelengths smaller than 5000 km. Reproduced from King SD (2002) Geoid and topography over subduction zones: The effect of phase transformations. *Journal of Geophysical Research-Solid Earth* 107.

Focusing on short-wavelength geoid (Figure 11), there is significant structure in the geoid that appears to be related to subduction zones but is obscured by the longer-wavelength components. Specifically, there is a narrow trough in the geoid (a local geoid minimum) that is approximately 200 km in width and spatially coincident with the trench (Figure 11). There are local geoid maxima in the back-arc region of almost every subduction zone in the short-wavelength geoid, and the geometry of the geoid maxima matches the geometry of the associated arcs. At this scale, there are geoid local maxima

associated with subduction zones even at locations where the association fails at the longer wavelengths (e.g., the Aleutians, Kamchatkan, and Middle America subduction zones).

The EGM96 geopotential model has been superseded by EGM2008 (Pavlis et al., 2012), which is complete to spherical harmonic degree and order 2159. For comparison, EGM96 was complete up to spherical harmonic degree and order 360. At the scales of interest for subduction modeling, there is no observable difference between the two models. One minor point that impacts geoid modeling is that the gravitational potential coefficients J_2 of a hydrostatic Earth model appear to be slightly smaller than the previous estimates (e.g., Nakiboglu, 1982) because of a correction to a secondary term in Nakiboglu's formulation and a revised estimate of the moment of inertia ratio (Chambat et al., 2010), both of which reduce the hydrostatic J_2 and thus increase the geoid. The difference between the polar and equatorial radii in the geoid increases from 98 to 113 ± 1 m.

The long-wavelength geoid associated with subduction zones (see **Section 7.09.2.6**) requires that subducting slabs encounter a resistance to flow, modeled as an increase in effective viscosity by a factor of 30, or more, at a depth of 670 km (Hager, 1984; Hager and Clayton, 1989; Hager and Richards, 1989; King and Hager, 1994; Ricard et al., 1989; Zhong and Gurnis, 1992). Because the equation for the gravitational potential is linear, we can decompose the problem; the total geoid anomaly is the sum of the contribution from each density anomaly within the Earth. This includes density anomalies that drive mantle convection, such as dense slabs or buoyant plumes, and density anomalies that result from deformed boundaries as a result of convection. This problem was first addressed by Pekeris (1935) but has been expanded on by others (e.g., McKenzie, 1977; Morgan, 1968; Ricard et al., 1984; Richards and Hager, 1984; Runcorn, 1967). Focusing on subduction zones for purposes of illustration, the mass excess due to the dense, sinking slab contributes a positive term to the total geoid while the mass deficit, due to the downwarping of the surface above the slab, contributes a negative term to the total geoid. These contributions are similar in magnitude and opposite in sign. The resulting total geoid is sensitive to variations in either the density structure of the slab or the deformation of the surface. The magnitude of the downwarping of the surface is a strong function of the viscosity structure of the medium (cf. Ricard et al., 1984; Richards and Hager, 1984). (It should be noted that the boundary between the mantle and core is also deformed as a result of the flow driven by the sinking slab; however, at the wavelengths appropriate to subduction zone problems, this is not a significant contribution even though it is included in most geoid calculations, Richards and Hager, 1984.)

While the density structure of a subducting slab is a complex function of temperature, chemical variations, and phase transformations, there is little doubt that subducting slabs are denser than the surrounding mantle. In order to create a positive geoid anomaly over a subducting slab in an isoviscous fluid, it would require that estimates of slab thermal structure underpredict the density anomaly due to the slab by approximately a factor of 4 (Davies, 1981) or that approximately 300 km of slab material is piled in the transition zone (Hager, 1984). This is much greater than the combined uncertainties in the thermal structure, thermal properties of the slab, and the errors made by neglecting the compositional component of the slab. Thus, some form of resistance to the downward motion of the slab is necessary to reduce the associated surface deformation. In the majority of studies on subduction, it has been assumed that the resistance to the slab is provided by an increase in viscosity at some depth (usually taken to be between the upper mantle and lower mantle). King (2002) had shown that the phase transformation from ringwoodite to perovskite plus ferropericlase is insufficient to change the sign of the geoid. With a sufficient reduction in the surface deformation, provided by the increase in viscosity of the lower mantle, the total geoid anomaly over the downwelling slab can be positive. In this way, the association of local maxima in the geoid with subduction zones provides a constraint on the viscosity of the mantle.

Based on the analysis described earlier in the text, an increase in mantle viscosity with depth by a factor of 30–100 is required to produce local geoid maxima over subducting slabs. The low end of this range is consistent with studies of other geophysical observations that constrain mantle rheology including postglacial rebound (cf. Forte and Mitrovica, 2001; Mitrovica, 1996; Mitrovica and Forte, 1997), plate velocities (cf. Lithgow-Bertelloni and Richards, 1995), and rotation. The high end of the viscosity increase required by the subduction zone geoid observation is more difficult to reconcile with most mantle viscosity models from postglacial rebound studies, although a few postglacial rebound studies have predicted a viscosity increase this large (e.g., Lambeck et al., 1990). It is also possible that both observations are compatible because the depths to which the postglacial rebound observation and the subduction zone observation are sensitive are different. For example, some models of mantle viscosity predict an increasing viscosity with depth throughout much of the lower mantle (e.g., Ricard and Wuming, 1991). If these viscosity models are correct, then it is possible to reconcile a higher value of lower mantle viscosity predicted by the subduction zone observation because the postglacial rebound observation is sensitive to the rheology in the upper 1500 km (Mitrovica, 1996) while the geoid is more sensitive to the lower mantle viscosity.

7.09.3 Kinematic Models for Slab Thermal Structure

Slab thermal modeling takes a complex region of the Earth where the downgoing plate descends into the mantle and attempts to capture this in a mathematically tractable manner. The realization that cold oceanic plate was sinking into the mantle (e.g., Venig Meines, 1954) quickly led to the development of thermal models for the subducting slab and the overlying plate (Mckenzie, 1969; Minear and Toksöz, 1970) (**Figure 12**). The first analytic thermal model for the subducting slabs was presented by McKenzie (1969). The thermal effect of mantle wedge flow was incorporated with an analytic corner flow solution. Using these kinematic models, the effect of frictional heating along the interface was explored. The results were used in petrologic, seismic, or mineral physics investigations (e.g., Bina, 1996, 1997; Helffrich et al., 1989; Kirby et al., 1996; Marton et al., 1999, 2005; Minear and Toksöz, 1970; Molnar et al., 1979; Peacock, 1991, 1996,

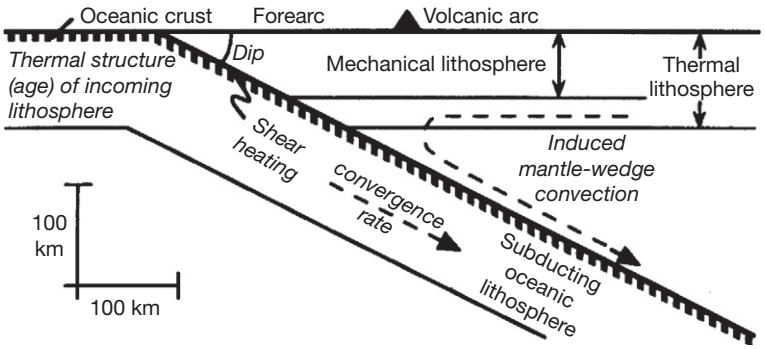


Figure 12 Cartoon cross section of a subduction zone showing the major features in slab thermal models. Reproduced from Peacock S (1996) Thermal and petrologic structure of subduction zones. In: Bebout GE, Scholl D, Kirby S, and Platt J (eds.) *Subduction: Top to Bottom*. pp. 119–133. Washington, DC: American Geophysical Union.

2003; Peacock et al., 1994; Staudigel and King, 1992). In this section, we first describe the McKenzie model and a corner flow solution and then discuss what we have learned from the modeling results about the thermal effect of frictional heating and the petrologic evolution of subducting slabs.

7.09.3.1 McKenzie Model

The first analytic thermal model for the subducting slab, presented by McKenzie (1969), assumes a hot, uniform temperature mantle surrounding the slab and solves the equation for the conservation of energy (see **Chapter 7.02**):

$$\rho c_p\left(\frac{\partial T}{\partial t}+v\cdot\nabla T\right)=\nabla\cdot(k\nabla T)+H \qquad [3]$$

where ρ is the density of the slab, c_p is the coefficient of specific heat, T is the temperature, v is the slab velocity, k is the thermal conductivity, and H is the rate of internal heat generation. For steady-state heat transfer, eqn [3] reduces to

$$\rho c_p(v\cdot\nabla T)=\nabla\cdot(k\nabla T)+H \qquad [4]$$

The term on the left-hand side accounts for advective heat transfer, and the first term on the right-hand side represents conductive heat transfer.

Substituting characteristic values for the temperature difference between the mantle and slab, T_0, and the length scale of the problem, l (e.g., the width of the slab) (see **Chapter 7.04**),

$$T=T_0T';\ x=lx';\ z=lz' \qquad [5]$$

then aligning the coordinate system such that the x-axis is along the downdip slab direction and the z-axis is perpendicular, we can write eqn [4] as

$$\frac{\partial^2T'}{\partial x'^2}-2R\frac{\partial T'}{\partial x'}+\frac{\partial^2T'}{\partial z'^2}=0 \qquad [6]$$

where R is the thermal Reynolds number, defined as $R=\frac{\rho c_p v_x l}{2k}$. Note that v becomes v_x in this coordinate system and v_z is zero because the assumption is that the slab is two-dimensional and does not deform. If we assume that the temperature at the top and temperature at the bottom of the slab are maintained at a uniform temperature by the surrounding mantle, the solution can be simplified to

$$T'\approx 1-\frac{2}{\pi}\exp\left(-\frac{\pi^2x'}{2R}\right)\sin\pi z' \qquad [7]$$

There are several generalizations of kinematic slab thermal models that are illustrated clearly by this derivation. First, the mantle surrounding the slab is assumed to be at a uniform temperature and is not locally impacted by the slab. Second, the slab is assumed to descend at a uniform rate that is always parallel to the dip of the slab everywhere (e.g., no slab buckling or internal strain). Finally, the slab does not exchange mass with the surrounding mantle (e.g., no thermal erosion, no melting, or dehydration).

Using this model, McKenzie concluded that the maximum depth of the deepest earthquakes in the Tonga–Fiji–Kermadec arc coincides with the depth of the 680 °C isotherm in the model. While he avoided attaching a significance to the exact value of the isotherm, noting the simplicity of the calculations, he suggested a thermal control to the maximum depth extent of deep earthquakes, a suggestion that has been subsequently questioned (e.g., Chen et al., 2004). In addition, he used the calculated thermal structure to quantify a buoyancy force due to the cold subducting slab that he estimated at $f=2.5$ sin ϕ kbars, where ϕ is slab dip.

Some of the limiting assumptions in the McKenzie slab model are addressed in the numerical model by Minear and Toksöz (1970). In the Minear and Toksöz model, the slab material moves with a fixed velocity relative to a stationary mantle, and the computational scheme consists of translating the temperature field along the slab dip direction and allowing the slab to equilibrate over a fixed time interval. As they stated in the text, "in essence, we have assumed the dynamics and have computed the temperature field given this motion field." An appealing aspect of the Minear and Toksöz model is that the slab geometry, age of the incoming ocean lithosphere, and slab velocity are specified, making it a fairly straightforward exercise to model observed subduction zone geometries. In addition, the effect of additional sources of energy such as adiabatic compression, phase transformations, and shear heating could be incorporated in the slab thermal structure through the imposed heat source term, H, in eqn [3]. Based on their resolution analysis, Minear and Toksöz (1970) estimated that the temperatures resulting from these computations 'are probably in error by less than 10%.' Because this model enables researchers to calculate the thermal structure of a subducting

slab using geophysical observable constraints such as slab dip, incoming plate age, and plate velocity, the Minear and Toksöz model has been used extensively (e.g., Bina, 1996, 1997; Helffrich et al., 1989; Ita and Stixrude, 1992; Kirby et al., 1996; Marton et al., 1999, 2005; Spiegelman and Mckenzie, 1987; Stein and Stein, 1996; Wiens, 2001).

7.09.3.2 Corner Flow Solution

Another kinematic approach to the subduction problem is to impose the mantle flow field calculated by the corner flow solution (e.g., Batchelor, 1967; Mckenzie, 1969; Peacock et al., 1994; Stevenson and Turner, 1977; Turcotte and Schubert, 1982) and solve for the thermal field of the subducting slab and the mantle wedge. This is sometimes referred to as the 'Batchelor solution' after the original reference (i.e., Batchelor, 1967). With the corner flow solution, the velocity field of the slab and the surrounding mantle, including the mantle wedge, satisfy the equations of motion for a constant viscosity fluid. This solution uses the stream-function formulation (see **Chapter 7.04**) to solve the equations of motion assuming a uniform velocity, rigid slab, and plate. It is possible to write the stream function, ψ, in the form

$$\psi = Ax + Bz + (Cx + Dz)\arctan(zx) \quad [8]$$

where A, B, C, and D are constants to be determined by the boundary conditions. The x-axis is in the horizontal direction and the z-axis is in the vertical axis. Because of the geometry of the descending slab, there are two distinct stream functions, that is, two sets of constants to be solved for in this problem: one above and one below the slab. The general problem leads to quite complicated integrals; however, a slab dip of $\pi/4$ allows easy evaluation of the integrals.

Solving for the flow in the mantle wedge, we have the following boundary conditions:

$$u = v = 0 \text{ on } z = 0,\ x > 0 \text{ or } \arctan\frac{z}{x} = 0 \quad [9]$$

$$u = v = \frac{U\sqrt{2}}{2} \text{ on } z = x \text{ or } \arctan\frac{z}{x} = \frac{\pi}{4} \quad [10]$$

where U is the imposed plate velocity and u and v are the horizontal and vertical components of the flow velocity. By straightforward substitution, it is possible to show that the pressure P in the arc corner is

$$P = \frac{\eta U\sqrt{2}[\pi x + (4-\pi)z]}{(2-\pi^2/4)(x^2+z^2)} \quad [11]$$

where η is the viscosity of the fluid. We can evaluate this expression using the fact that $x = z = r\sqrt{2}/2$ where r is the length along the slab and we find that

$$P = \frac{4\eta U}{(2-\pi^2/4)r} \approx \frac{-8\eta U}{r} \quad [12]$$

The negative pressure means the back-arc region flow tends to lift the slab against the force of gravity. Notice that the pressure varies with $1/r$ along the slab, with a singularity at the trench. The total lifting torque on the slab will be $r \times P$. For the ocean side of the trench, the boundary conditions are

$$u = U,\ v = 0 \text{ on } z = 0,\ x > 0 \text{ or } \arctan\frac{z}{x} = \pi \quad [13]$$

$$u = v = \frac{U\sqrt{2}}{2} \text{ on } z = x \text{ or } \arctan\frac{z}{x} = \frac{\pi}{4} \quad [14]$$

Again, by substitution, the pressure on the bottom side of the slab is given by

$$P = \frac{\eta U}{r}\left(\frac{3\pi\sqrt{2}-4}{9\pi^2/4-2}\right) \approx \frac{0.5\eta U}{r} \quad [15]$$

Given the geometry and the signs of the pressures, this flow also exerts a lifting torque on the slab that is about a factor of 20 smaller than the torque on the slab from the arc flow (Stevenson and Turner, 1977).

The corner flow solution does not overcome the rigid slab approximation, nor does it allow for a slab velocity that is consistent with the slab driving force. While the stream-function solution solves the equations of motion, a solution to the energy equation is also needed and several different solution strategies have been employed (cf. Peacock, 1996). Because of the discontinuity in the velocity boundary conditions at the corner of the wedge, there is a singularity in the pressure field that presents problems for fluid/melt flow models (e.g., Spiegelman and McKenzie, 1987).

Figure 13 shows the results of two different kinematic slab thermal models: one based on the method described in Minear and Toksöz (1970) and the other based on the analytic corner flow solution described in the preceding text plus the streamline upwind/Petrov–Galerkin solver from ConMan (King et al., 1990) for the temperature (energy) equation. The solution is for a 50-Myr-old slab with dip angle of 45° and subduction rate of 5 cm year^{-1}. The complete list of parameters for these calculations is provided in **Table 1**. There are two reasons for the significant difference in the slab thermal structure between the two solutions. First, in the Minear and Toksöz (1970) thermal model, the slab curves (bends) as it enters the subduction zone, whereas in the corner flow solution, the slab dip is uniform at 45° throughout the domain. This accounts for the difference between the two solutions seen in the top 50 km depth. Second, the Minear and Toksöz (1970) thermal model does not include the effect of the mantle wedge flow on the slab thermal structure. In the Minear and Toksöz (1970) model, perpendicular to the slab, heat is transferred by conduction only. We used the induced flow approximation that is included in the Minear and Toksöz code, but this does not improve the agreement between the solutions. This accounts for the majority of the difference in the region deeper than 50 km depth. The effect of the induced wedge flow on the slab thermal structure becomes even more significant when a temperature-dependent mantle rheology is applied (see **Section 7.09.4.2**). Returning to the Minear and Toksöz claim that the solution is in error by less than 10% (due to the grid), it is important to recognize that this is a statement of the numerical solution on the grid that they used in their paper relative to the solution with the same basic underlying physical assumptions in their model on a more highly refined grid. The difference between the Minear and Toksöz solution and the Batchelor solution plus advection is much more than 10% for the problem described in the preceding text (**Figure 14**).

In the ConMan solution, we use grid spacing on the order of 1 km in the corner of the wedge. We find that grid spacing

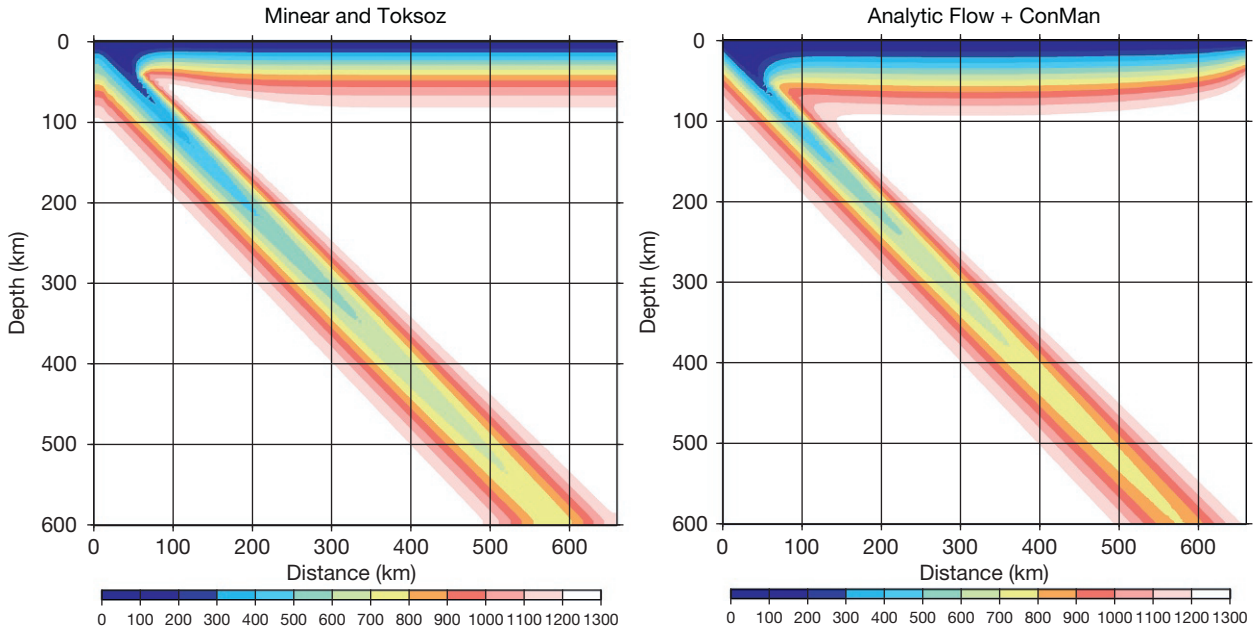


Figure 13 A comparison of slab thermal structures from a 50-Myr-old, 45°-dipping slab descending at 5 cm year^{-1} with kinematic slab approximations (left) Minear and Toksöz (1970) and (right) the analytic corner flow solution for flow and ConMan's temperature solver. Below 50 km depth, the Minear and Toksöz slab is colder than the corner flow plus ConMan solution.

Table 1 Parameters used in subduction zone thermal models shown in Figures 7 and 8

Parameter	*Symbol*	*Value*
Slab velocity	v	5 cm year^{-1}
Slab dip	Θ	45°
Box depth	d	600 km
Box width	w	660 km
Slab age	T_{slab}	50 Myr
Thickness of overriding plate	d_{plate}	50 km
Density	ρ	3300 kg m^{-3}
Thermal conductivity	k	3.0 W m^{-1} K^{-1}
Specific heat	c_p	1250 J kg^{-1} K^{-1}
Thermal diffusivity	$\kappa = \frac{k}{\rho c_p}$	7.272×10^{-5} m^2 s^{-1}
Reference temperature	T_0	1300 °C

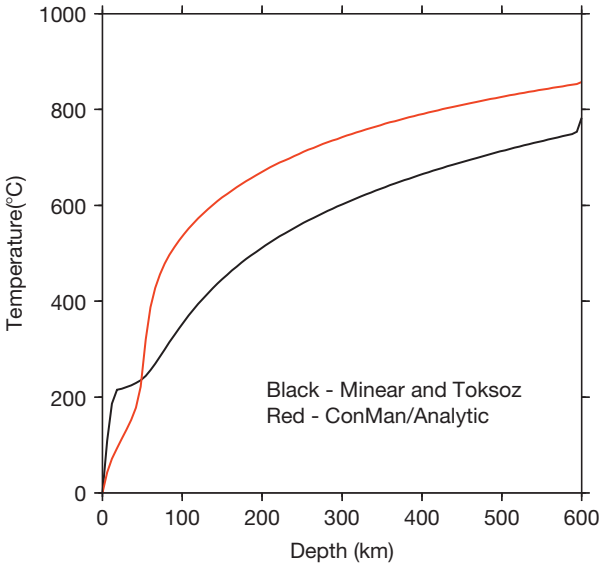


Figure 14 Temperature along the line $x=z$ (i.e., the slab interface) for the kinematic slab thermal models shown in the previously mentioned text. The difference of approximately 150 °C is largely due to the lack of corner flow in the Minear and Toksöz model.

on this order is necessary to resolve the flow near the pressure singularity near the corner. In a community benchmarking study for subduction zone thermal models by van Keken et al. (2008), some contributors used even finer mesh, reaching grid resolution of a few to tens of meters. In the real Earth, it is likely that because of the effects of temperature-dependent rheology and serpentinization near the corner of the wedge, flow velocity increases downdip more gradually. Thus, the singularity at the wedge tip present in the corner flow solution does not occur in reality.

7.09.3.3 Earlier Estimates of Frictional Heating Along the Plate Interface

Subduction zones are where cool surface material is advected into the mantle and is therefore cold relative to the average upper mantle temperature, yet they are the sites of active arc volcanism. The early postplate tectonic view is that island arc volcanism was the result of slab melting (e.g., Ringwood, 1975) that required temperatures approaching the melting point of basalt. One possible mechanism for elevating the slab surface temperature was large rates of frictional heating along the plate interface (e.g., Oxburgh and Turcotte, 1968). Temperature conditions along the shallow part of the interface

are also important to such processes as dehydration reactions and phase changes of subducted sediments and crust, subduction thrust earthquakes, and ETS events. To quantify shear heating along the plate interface, Molnar and England (1990) developed analytic expressions for the steady-state temperature in the top 50 km of the subduction zone. Their expressions are

$$T = \frac{(Q_0 + Q_{SH})z_f/k}{S} \quad [16]$$

where Q_0 is the basal heat flux (W m^{-2}), Q_{SH} is the rate of shear heating (W m^{-2}), z_f is the depth to the fault (m), k is the thermal conductivity (W m^{-1} K^{-1}), and S is given by

$$S = 1 + b\sqrt{vz_f \sin\phi/\kappa} \quad [17]$$

where b is a constant of order 1 (in the case of Figure 15, $b = 10$), v is the convergence velocity, ϕ is the angle of subduction, and κ is the thermal diffusivity (m^2 s^{-1}). The slab surface temperature calculated by using this expression with values from Table 1 and Q_0 as 35 mW m^{-2} is shown in Figure 15.

A series of calculations described by Peacock et al. (1994) and Peacock (1996) solve the energy equation (eqn [3]), using a finite-difference formulation described in Peacock (1991) and the Batchelor's corner flow solution (Batchelor, 1967; see Section 7.09.3.2). In the study by Peacock et al. (1994), a range of shear heating rate was tested using (1) frictional heating with a constant τ, (2) frictional heating with pressure-dependent τ with variable μ' values, or (3) frictional heating above a temperature-dependent brittle–ductile transition and viscous energy dissipation below the transition. The modeling results show that the uncertainty in the amount of shear heating, particularly below 50 km depth, leads to several hundred degrees difference in the temperature of the slab (Peacock, 1996). The results show that the assumed rate of shear heating has a profound effect on the slab thermal structure, with temperatures on the slab–wedge interface varying by several hundred degrees at 100 km depth (cf. Peacock, 1992, 1993, 1996; Peacock et al., 1994).

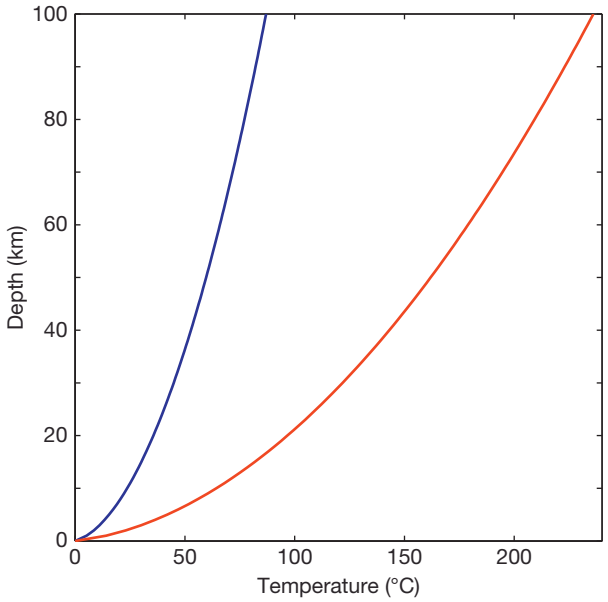


Figure 15 Temperature along the slab interface calculated with zero shear heating (blue) and with 60 mW m^{-2} (red), using eqns [16] and [17] and subduction parameters listed in Table 1.

Using surface heat flow observations, maximum depth of seismic coupling, and exhumed high-pressure, low-temperature metamorphic rocks as model constraints, Peacock (1996) concluded that with the exception of very rare circumstances, the slab does not melt; however, dehydration of Na-amphibole, lawsonite, and chlorite in the subducted oceanic crust can trigger hydrous melting in the overlying mantle wedge. As discussed in Section 7.09.2.3, more recent petrologic and geochemical data indicate that the melting of the subducted sediments and oceanic crust occurs commonly. We will discuss more recent modeling results for the effect of frictional heating and other possible mechanisms raised for melt generation in subduction zones in Section 7.09.4.

Kinematic slab models were used to map the mineralogy of the slab at deeper depths (e.g., Bina, 1996, 1997; Ita and Stixrude, 1992; Marton et al., 1999, 2005). While kinematic models have provided an estimate of where phase transitions might occur, they do not have the predictive ability to follow the implications of the mineralogical models as the slabs evolve dynamically (e.g., Ita and King, 1998). For example, these mineralogical models predict significant changes in slab density as the slab subducts, and the changes are expected to impact the plate and slab velocities (e.g., Ita and King, 1998; King and Ita, 1995), but this effect cannot be accounted for in the kinematic models. To model the dynamic effect of phase transitions, such as that of metastable olivine, larger-scale dynamic models are used (e.g., Christensen, 1996; Gurnis and Hager, 1988; Ita and King, 1998; Zhong and Gurnis, 1994, 1997). Dynamic models are discussed in Section 7.09.5.

7.09.4 Coupled Kinematic–Dynamic Slab–Wedge Models

For regional-scale, isolated slab models, a kinematic description of slab motion remains popular in part because the input parameters are easily observed properties of subduction zones (i.e., plate age, plate velocity, and slab dip), yet the fully kinematic slab–mantle wedge model with the analytic corner flow solution is inflexible with the slab geometry (e.g., curved geometry) and mantle rheology (e.g., temperature-dependent rheology as opposed to isoviscous rheology used in the corner flow solution), and the dynamics of mantle wedge flow cannot be explored with kinematic models. Dynamic subduction models, as will be discussed in Section 7.09.5, allow the incorporation of the dynamic effects of the mantle flow and the subducting slab, though the inability to control the model geometry and subduction parameters, such as subduction rate, makes it difficult to study the thermal and flow fields in specific subduction zones, and the large-scale nature of dynamic models (to encompass all the forces involved in the dynamic system) makes it difficult to achieve the resolution needed for regional-scale studies. Coupled kinematic–dynamic calculations (e.g., Arcay et al., 2005; Conder et al., 2002; Currie et al., 2004b; Davies and Stevenson, 1992; Furukawa, 1993; Honda and Saito, 2003; Kelemen et al., 2003; Manea et al., 2005; Peacock et al., 2005; Peacock and Wang, 1999; Syracuse

et al., 2010; van Keken et al., 2002; Wada and Wang, 2009) have therefore become a popular approach, for example, to investigate the effects of slab geometry and mantle rheologies on the mantle wedge flow pattern and the slab–wedge thermal structure; to predict the effect of the slab and wedge flow on surface heat flow; to derive petrologic structures; to quantify fluid flux from the slab; to evaluate the dependence of seismicity on temperature, metamorphic state, and fluid distribution; to model fluid migration in the overlying mantle wedge; and to estimate subarc mantle temperature and melting location.

In this section, we provide a brief description of 2-D coupled kinematic–dynamic models and discuss viscous coupling between the subducting slab and the overlying mantle, which drives mantle wedge flow, and the role of temperature dependence of the mantle wedge rheology. Using the reported modeling results, we discuss the implications of the predicted thermal and petrologic structures to the metamorphic state and seismicity in the slab, fluid release from the slab, and melt generation in subduction zones. Finally, we discuss the effect of slab geometry on the mantle flow and the thermal structure in 3-D.

7.09.4.1 Governing Equations and Boundary Conditions

Coupled kinematic–dynamic models for the slab–mantle wedge system typically consist of a nondeforming overlying plate, a viscous mantle wedge, and a subducting slab with prescribed motion (**Figure 16(a)**) (e.g., Arcay et al., 2005; Conder et al., 2002; Currie et al., 2004b; Davies and Stevenson, 1992; Furukawa, 1993; Honda and Saito, 2003; Kelemen et al., 2003; Manea et al., 2005; Peacock et al., 2005; Peacock and Wang, 1999; Syracuse et al., 2010; van Keken et al., 2002; Wada and Wang, 2009). The geotherm along the trench-side vertical boundary is prescribed, as in the case for kinematic models (**Section 7.09.3**). The geotherm along the back-arc-side vertical boundary is prescribed from the surface down to the depth of an inflow–outflow transition of the mantle wedge, which is determined through the model calculation. The back-arc geotherm is commonly calculated to match observed or typical back-arc surface heat flow for the shallow part and by using a mantle potential temperature and adiabatic gradient for the deeper part. With the boundary conditions and prescribed material properties, including mantle rheology, the steady-state thermal and flow fields are calculated by solving the energy equation (eqn [3] for the Boussinesq approximation (BA) or a dimensional steady-state form of eqn [46] for the extended Boussinesq approximation (EBA)) (see **Section 7.09.5.1** for different approximations) and the equations for the conservation of mass and momentum for incompressible viscous flow:

$$\nabla\cdot\vec{u}=0 \qquad [18]$$

$$-\nabla P+\nabla\cdot\sigma'+\rho' g=0 \qquad [19]$$

where u is the flow velocity, σ' is the deviatoric stress tensor, P is the dynamic pressure (i.e., the nonhydrostatic portion of pressure that is generated by flow), and ρ' is the incremental density variation defined as

$$\rho'=-\rho_0\alpha_v(T-T_0) \qquad [20]$$

where ρ_0 is the reference density at a reference temperature T_0 and α_v is the coefficient of thermal expansivity at constant volume (see **Chapter 7.02**). The buoyancy (third) term $\rho' g$ in eqn [21], however, is commonly omitted, though there are models that specifically address the effect of thermal and chemical buoyancy exerted by free fluids and partially molten diapirs in the mantle wedge (e.g., Gerya, 2011; Gerya et al., 2006; Gorczyk et al., 2006; Hebert et al., 2009; Zhu et al.,

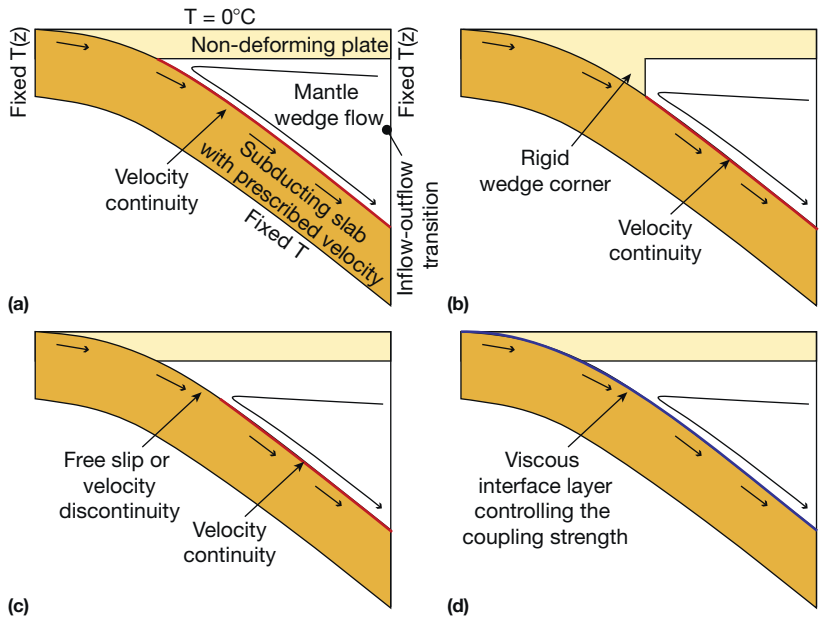


Figure 16 Schematic diagrams showing (a) typical boundary conditions used in coupled kinematic–dynamic models for subduction zones and different approaches to satisfy low surface heat flow in the forearc: (b) a rigid corner approach, (c) a free-slip or velocity discontinuity approach, and (d) an interface layer approach.

2011). A time-dependent thermal field has also been solved with kinematic–dynamic models; the initial thermal condition is assigned using a oceanic or continental geotherm, and the change in the thermal field with time due to the imposed motion of the slab with time is modeled using eqn [4] as opposed to eqn [3] (e.g., Gerya et al., 2006; Gorczyk et al., 2007, 2006; Spinelli and Wang, 2008; Wang et al., 1995). The heat generation term H in eqn [3] may include radiogenic heat production (Q_{RH}), frictional heating (Q_{FH}), and heat generation by viscous energy dissipation (Q_{VD}). Latent heat of hydration and dehydration reactions in the subducting slab and the mantle wedge depends on the distribution of hydrous phases in the subducting slab and the degree of hydration in the mantle wedge (Peacock, 1996). On the basis of typical estimates for the H_2O content in the subducting slab and the mantle wedge, the thermal effects of these reactions are expected to be relatively small (Peacock, 1996, references therein) and are commonly neglected (latent heat is included in Gerya et al., 2006; Gorczyk et al., 2006; Hebert et al., 2009). The effect of neglecting the mantle compressibility was investigated by Lee and King (2009); their study indicates that this effect is negligibly small. Benchmarking results of seven different numerical codes for this type of subduction zone thermal modeling are reported in van Keken et al. (2008).

The hybrid subduction zone models simulate slab-driven mantle flow only, and the effect of thermal buoyancy is commonly neglected, assuming that the effect of buoyancy in the forearc and arc regions is likely negligible. This assumption may be reasonable because of the narrow space and relatively small temperature difference between the top and bottom of the mantle wedge that drives buoyancy flow. However, buoyancy in the mantle wedge may be dynamically important if the viscosity in the wedge is significantly reduced due to relatively high temperature, high volatile content, and high strain rates (e.g., Kelemen et al., 2003). Modeling the effect of buoyancy-driven mantle flow in the hybrid model remains a challenge due to the nature of the stress-free boundary conditions at the wedge boundary and the stability of the overriding plate.

7.09.4.2 Slab–Mantle Viscous Coupling

The trenchward extent of wedge flow, which has a strong effect on the thermal structure of the subducting slab, depends on the depth at which the slab and the overlying mantle become viscously coupled. Earlier studies that compared the thermal modeling results with surface heat flow observations (e.g., Furukawa, 1993; Honda, 1985) showed that with a conventional corner flow solution or with viscous coupling over the entire length of the slab–mantle interface, the thermal model overpredicts surface heat flow in the forearc region (**Figure 16(a)**). To satisfy the observed heat flow, various approaches have been taken, such as imposing a rigid corner to prevent the hot mantle from entering the wedge corer (e.g., Currie et al., 2004b; Peacock and Wang, 1999; van Keken et al., 2002) (**Figure 16(b)**), applying free slip or velocity discontinuity to the shallow part of the slab–mantle interface (e.g., Furukawa, 1993; Kelemen et al., 2003; Kneller et al., 2005, 2007) (**Figure 16(c)**), applying a thin interface layer of known strength to simulate various degree of slab–mantle viscous coupling (e.g., Wada et al., 2008) (**Figure 16(d)**), and applying weak nodes along the shallow part of the slab–mantle interface to a temperature- and strain rate-defined brittle–ductile transition depth, below which velocity continuity is applied (e.g., Conder, 2005). Depending on the depth of decoupling, type of weak zone, and slipping fault parameters, the predicted thermal structure of the slab and the mantle varies significantly.

Based on a numerous comparison of these modeling results with geologic and geophysical observations, including surface heat flow, seismic attenuation and velocity structures, and petrologic and geomechanical constraints on subarc mantle temperature, Wada and Wang (2009) and Syracuse et al. (2010) found that the slab and the overlying mantle start to become fully coupled at 70–80 km depth for most if not all subduction zones. Further, in a model with the interface layer approach (**Figure 16(d)**), in which the degree of coupling is controlled by the rheologies of the interface and the overlying mantle, the shallow part of the wedge is nearly completely decoupled due to a strong wedge corner that results in shear strain localization within the relatively weak interface layer, and the transition from decoupling to coupling is sharp due to nonlinear feedback between rheology, temperature, and mantle flow. This result indicates that other decoupling approaches mentioned in the preceding text are reasonable approximations since they impose a sharp change in flow velocity in the wedge at a prescribed depth. The sharp decoupling–coupling transition leads to a sharp thermal transition within the wedge, which is consistent with the sharp change in seismic attenuation (Wada et al., 2011), and it also results in a sharp downdip increase in temperature along the slab surface by up to ~200 °C near the transition. We will revisit the effect of the decoupling–coupling transition on the thermal structure of the subducting slab and the mantle wedge in **Sections 7.09.4.4** and **7.09.4.5**.

One of the outstanding questions in the study of slab–mantle wedge dynamics is the mechanism that controls the decoupling–coupling transition. Decoupling occurs when the interface is weaker than the overlying mantle. Therefore, the key parameter is the strength contrast between the interface and the overlying mantle (Wada et al., 2008). The shallow decoupling is often attributed to the presence of weak hydrous minerals and high pore fluid pressure along the interface, but what controls the decoupling to disappear at 70–80 km depth? Something is causing the interface to strengthen or the overlying mantle to weaken. Wada and Wang (2009) raised several potentially important factors to the transition depth, such as fluid release from the subducting slab, dehydration of hydrous minerals along the interface, and hydration of mantle minerals at the base of the mantle wedge, but these processes occur over a wide depth range with a significant variability among different subduction zones, and therefore, they are unlikely to be the key mechanisms that control the relatively uniform decoupling–coupling transition depth. Deformation experiments on antigorite and olivine indicate that their strength contrast disappears as pressure and temperature increases (Hilairet et al., 2007), which may lead to the decoupling–coupling transition. Wada and Wang (2009) also suggested that the thermal state of the back-arc may play a key role in controlling the transition by controlling the amount of heat transferred into the mantle wedge.

7.09.4.3 Isoviscous to Temperature-Dependent Mantle Rheologies

Earlier models assume an isoviscous mantle rheology as in the case with the corner flow solution. Figure 17(a) shows the temperature field predicted by a model with an isoviscous wedge using the convection code ConMan (King et al., 1990). The mesh is nonuniform with significant refinement near the corner of the wedge with a grid spacing of approximately 1 km. The grid extends to 660 km × 600 km and the boundary conditions along the inflow/outflow sides of the wedge are natural boundary conditions (i.e., normal and tangential components of total stress are zero). In the model, the slab descends at 45° at a fixed velocity of 5 cm year^{-1}. The incoming plate material is 50 million years old, and the velocity of the overriding plate is set to zero to a depth of 50 km. From the surface to 50 km depth, a free velocity discontinuity is imposed along the interface (Figure 16(c)), and downdip of 50 km depth, velocity continuity across the interface simulates the viscous coupling between the subducting slab and overlying mantle that drives mantle wedge flow. The temperature contours in the mantle wedge (Figure 17(a)) and the stream lines of the mantle wedge flow (not shown in Figure 17(a)) are relatively evenly distributed because of the isoviscous rheology, and the resultant flow pattern is similar to that of the analytic corner flow solution.

Peacock and Wang (1999) developed a 2-D finite-element thermal model with an isoviscous mantle wedge and the rigid corner (Figure 16(b)) for two end-member warm-slab and cold-slab subduction zones, Nankai and NE Japan, respectively. Their models show evenly distributed mantle flow similar to the corner flow solution, except for the effects of slab curvature. They also showed the strong contrast in the thermal structures of the two subduction zones.

At the temperature and pressure conditions in the upper mantle, Earth materials deform in a ductile manner primarily through diffusion and dislocation creep, and the strong temperature dependence of the viscosity of the mantle had been known from deformation experiments (e.g., Hirth and Kohlstedt, 2003; Karato and Wu, 1993) (see Karato, 2011 for the historical review of the rheology of the Earth's mantle). The transition between diffusion creep and dislocation creep depends on various factors, such as shear stress, water content, and thermal gradient, but in general, dislocation creep is thought to dominate in the shallow (<~200 km depth) asthenospheric mantle (Karato and Wu, 1993). This is supported by the seismic anisotropy observed in the upper mantle (Savage, 1999), which is thought to result from the reorientation of mineral grains, particularly olivine, during dislocation creep (Section 7.09.2.4).

The constitutive relation between the deviatoric stress tensor (σ_{ij}) and the strain rate tensor ($\dot{\varepsilon}_{ij}$) for diffusion-creep rheology is

$$\sigma_{ij} = 2\eta\dot{\varepsilon}_{ij} \quad [21]$$

where η is the viscosity of the fluid. The relation for the dislocation-creep rheology can be expressed in the same form as eqn [21] but using the effective viscosity η_e:

$$\sigma_{ij} = 2\eta_e\dot{\varepsilon}_{ij} \quad [22]$$

The general form of the flow law for diffusion creep and dislocation creep is

$$\dot{\varepsilon} = Ad^{-p}\sigma^n C_{OH}^r \exp\left[-\frac{E+PV}{RT}\right] \quad [23]$$

where A is the preexponential factor, p is grain-size exponent, n is the stress exponent, C_{OH} represents water content and r the water exponent, E is the activation energy, V is the activation volume, P is the confining pressure, R is the universal gas constant (8.3145 J mol^{-1} K^{-1}), and T is the absolute temperature (e.g., Hirth and Kohlstedt, 2003; Karato and Wu, 1993) (see Chapters 2.18 and 7.02). In both diffusion creep and dislocation creep, the strain rate is highly sensitive to temperature. In the diffusion creep, the strain rate increases linearly with stress, following the Newtonian flow law ($n=1$), but decreases with grain size ($p>0$). In the dislocation creep, the strain rate increases nonlinearly with stress, following a power

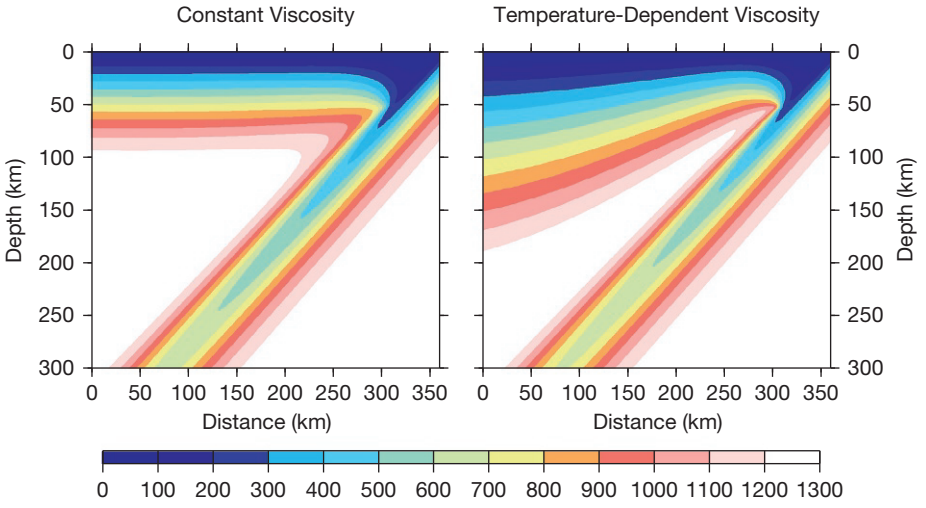


Figure 17 Thermal structure from two slab calculations with (left) a constant viscosity mantle wedge and (b) a diffusion-creep (Newtonian) mantle wedge, using ConMan (King et al., 1990) and subduction parameters listed in Table 1.

law ($n > 1$), but is independent of grain size ($p = 0$). In the relatively shallow part of the subduction zones, the pressure term PV is very small relative to E and is commonly neglected in the kinematic–dynamic subduction zone models (but not in the larger-scale dynamic models). By substituting eqn [21] or [22] into [23], one can obtain η in the case of diffusion creep or η_e in the case of dislocation creep, respectively.

Figure 17(b) shows the temperature field calculated by the same model as in **Figure 17(a)** except that rheological parameters for diffusion creep are assumed (a preexponential factor A' of 1.32×10^9 Pa s and an activation energy of 335 kJ mol^{-1}) following the community benchmarking study of van Keken et al. (2008), using the form

$$\eta = A' \exp\left[-\frac{E}{RT}\right] \qquad [24]$$

One of the important effects of a temperature-dependent rheology is that a thermal boundary layer develops in the shallow part of the wedge due to low temperature, and flow into the wedge corner is focused, bringing hot material closer to the wedge tip, compared with the isoviscous corner flow model. This focused flow leads to higher temperatures in the wedge corner and causes a sharp increase in the temperature along the slab–wedge interface near the wedge tip (e.g., Arcay et al., 2005; Currie et al., 2004b; Furukawa, 1993; Kelemen et al., 2003; Manea et al., 2005; Peacock et al., 2005; van Keken et al., 2002). The heat flow near the trench is elevated by 10% in the Newtonian rheology case, compared with the isoviscous calculation. The temperature in the downgoing slab differs by 98 °C at the slab–wedge interface at 150 km depth.

Furukawa (1993) applied a Newtonian rheology in his 2-D finite-difference model for NE Japan and compared the modeling results with surface heat flow observations. He showed that the change from lower heat flow in the forearc and high heat flow in the arc can be explained by slab–mantle decoupling to ~70 km depth, using the free-slip condition (**Figure 16(c)**). With the Newtonian mantle rheology, the flow in the mantle wedge is focused, resulting in higher temperature in the mantle wedge and at the slab surface beneath the arc than that predicted by a model with an isoviscous rheology.

Experimentally constrained rheological parameter values for olivine aggregates are available (e.g., Karato and Wu, 1993), including those that account for the effects of melts and aqueous fluids (e.g., Hirth and Kohlstedt, 2003). The common choice of parameters for the mantle wedge in the kinematic–dynamic model is those for dislocation creep of 'wet' olivine, in which water is incorporated as crystalline defects in nominally anhydrous olivine. This choice of rheology is based on geophysical observations, such as seismic anisotropy, that indicate dislocation creep in the upper mantle, the fact that olivine is the primary phase in the relatively high-temperature region of the upper mantle in which hydrous phases are no longer stable, and the assumption that water released from the subducting slab is distributed and incorporated as crystalline defects.

In kinematic–dynamic models with the BA, the effect of viscous energy dissipation in the mantle wedge is neglected (see **Section 7.09.5.1** for the BA). Studies that tested this effect show that in models with temperature-dependent mantle rheology, increased temperature due to viscous energy dissipation decreases the (effective) viscosity of the mantle. This feedback between the two parameters renders viscous energy dissipation self-defeating. Thus, Q_{VD} is negligibly small as concluded by many other authors (e.g., Currie, 2004). In a dynamic fully compressible convection formulation, viscous energy dissipation can be calculated consistently in the viscous flow formulation (e.g., Yuen et al., 1978). However, even in such a formulation, there are uncertainties because the deformation mechanism within the slab is poorly understood and, perhaps more importantly, because of the uncertainty and difficulty in representing the deformation between the slab and overriding plate.

In some studies that focus on processes in which the two deformation mechanisms (and others) play different roles, such as the development of mineral fabrics (e.g., Kneller et al., 2005) and the evolution of mineral grain size (e.g., Wada et al., 2011) in the mantle wedge, a composite rheology that accounts for both diffusion creep and dislocation creep is incorporated. The assumption that is made commonly in such case is that the two deformation mechanisms operate in parallel and the total strain rate (ε_{total}) is the sum of the two strain rates:

$$\varepsilon_{total} = \varepsilon_{diff} + \varepsilon_{disl} \qquad [25]$$

The mechanism that operates at a faster rate represents the dominant deformation mechanism. The effective viscosity (η_{eff}) is then described as

$$\eta_{eff} = \left(\frac{1}{\eta_{diff}} + \frac{1}{\eta_{disl}}\right)^{-1} \qquad [26]$$

The effect of the composite rheology on the overall mantle wedge flow pattern and the thermal structure is small based on the mantle wedge flow model of Wada et al. (2011) that adapts the grain-size evolution model of Austin and Evans (2007) based on the theories of static grain growth and dynamic recrystallization. The mantle wedge flow model shows that in the arc region of the mantle wedge, dislocation creep dominates the deformation except in the cold wedge nose. Another grain-size evolution model based on Zener pinning and a grain-damage theory, however, predicts that diffusion creep at high strain rate can lead to localized grain-size reduction and weakening, dominating the deformation (Bercovici and Ricard, 2012).

7.09.4.4 Slab Surface and Subarc Mantle Temperatures

There is a large degree of uncertainty in the model-predicted temperature along the slab surface. One of the factors that contribute to the uncertainty is the poorly constrained effect of shear heating, owing in part to large uncertainties in the magnitude of shear stress and in the downdip extent of the frictionally coupled zone of the interface. Compared to earlier estimates of shear stress along the interface of 100 MPa (e.g., Molnar and England, 1990; Scholz, 1990), mechanical models of forearc stresses constrained by focal mechanism of crustal earthquakes show low strength of subduction faults (Lamb, 2006; Seno, 2009; Wang and He, 1999; Wang and Suyehiro, 1999), and seismological and field observations (e.g., Magee and Zoback, 1993) and frictional heating studies (e.g., von Herzen et al., 2001; Wang et al., 1995) indicate that subduction

faults are weak and that the temporally and spatially averaged effective coefficient (μ') of friction of subduction faults is usually no greater than 0.05, typically around 0.03. The low μ' value of weak subduction faults results in low frictional heating (see Section 7.09.3.3 for μ' and frictional heating).

The updip and downdip extent of the frictionally coupled, seismogenic portion of the plate interface is controlled by changes in the sliding behavior of the material along the interface. Possible mechanisms for the changes in the sliding behavior have been discussed in the literature (e.g., Hyndman and Wang, 1993; Moore and Saffer, 2001; Saffer and Marone, 2003; Skarbek and Saffer, 2009; Tobin and Saffer, 2009; Vrolijk, 1990). The downdip extent of the seismogenic zone has been estimated for a number of subduction zones on the basis of the distribution of thrust events (e.g., Pacheco et al., 1993; Tichelaar and Ruff, 1993), geodetic data (e.g., Le Pichon et al., 1998; Simoes et al., 2004; Wallace et al., 2004; Wang et al., 2003), and tsunami modeling (e.g., Baba et al., 2002). The location of slow slip events may also help identify the downdip extent since they appear to occur in the transition zone between the seismogenic zone and the downdip aseismic section of the interface (Dragert et al., 2001; Obara et al., 2004; Ohta et al., 2006; Wallace et al., 2009). However, there are typically large uncertainties of several kilometers in these estimates.

At shallow depths, frictional coupling between the subducting slab and the overlying plate generates heat, increasing the temperature along and near the plate interface (Figure 18). Heat generation along the plate interface is illustrated in Figure 18. Frictional heating Q_{FH} is the product of shear stress τ and fault slip rate v, which is often approximated by convergence rate:

$$Q_{FH} = \tau v \qquad [27]$$

The shear stress on a fault at failure is

$$\tau = \mu(1-\lambda)\sigma_N = \mu'\sigma_N \qquad [28]$$

where μ is the coefficient of friction of the fault, μ' is the effective coefficient of friction, σ_N is the normal stress, and λ is the pore fluid pressure ratio,

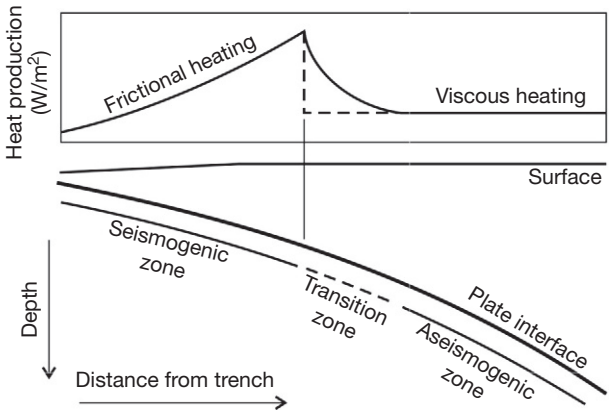


Figure 18 Schematic illustration of heat production along the plate interface.

$$\lambda = \frac{P_f}{P_l} \qquad [29]$$

where P_f is the fluid pressure and P_l is lithostatic pressure. A low μ' value of the fault results in low frictional heating. The effects of pore fluid pressure on the strength of subduction faults are discussed by Saffer and Tobin (2011).

Along the deeper, warmer part of the plate interface, below the brittle–ductile transition, shear heating occurs (Figure 18), and this is the product of shear stress and shear strain rate $\dot{\varepsilon}$:

$$Q_{VD} = \boldsymbol{\tau}\dot{\varepsilon} \qquad [30]$$

The effects of frictional heating and viscous energy dissipation along the plate interface are also investigated using kinematic–dynamic models, as will be discussed in Section 7.09.4. Viscous energy dissipation also occurs in the convecting mantle, which will be discussed further in Section 7.09.5.

Manea et al. (2005) investigated the Guerrero subduction zone in Mexico, characterized by flat subduction with a distant volcanic arc. Using a temperature-dependent rheology to model the temperature structure in the slab and wedge, they found significantly higher temperatures in the mantle wedge and at the slab surface below the volcanic arc compared with previous studies using an isoviscous mantle wedge (e.g., Currie et al., 2002). In their model, they included frictional heating to a depth of 40 km using Byerlee's friction law (Byerlee, 1978) with pore fluid pressure ratio of 0.98. In Manea et al.'s model, temperatures at the slab–wedge interface are also sufficient to allow for a slab contribution to the arc volcanism from the melting of hydrated sediment and oceanic crust, in agreement with geochemical observations of magmas in the Trans-Mexican Volcanic Belt Manea et al. (2005) also investigated the effect of buoyant melt in the wedge using the predicted wedge flow structures. Using the location of the Popocatepetl volcano, they estimated that magmatic blobs need to be larger than 10 km in diameter to have reasonable trajectories in the mantle wedge with realistic estimates for viscosity.

Peacock et al. (2005) constructed high-resolution 2-D finite-element models using the geometry along four profiles perpendicular to the Central American subduction zone in Nicaragua and Costa Rica. Realistic stress- and temperature-dependent rheology in the mantle wedge is necessary to have sufficiently high temperature in the mantle wedge below the volcanic arc to produce melting. Using the model, the effect of frictional heating along the interface was tested by assuming a constant shear stress. Frictional heating from the trench to a depth of 50 km with a constant shear stress of 40 MPa results in 58 °C higher slab surface temperature than the model without frictional heating. In their preferred model, they applied a lower constant shear stress of 10 MPa to a depth of 50 km. This subduction zone is characterized by rapid convergence of young oceanic lithosphere with minor changes in the convergence speed and age of the lithosphere along-strike but significant variations in the depth of the seismicity and the geochemistry of arc volcanism. The small along-strike variations in the convergence speed (increasing to the southwest from 79 to 88 mm year^{-1}) and age (decreasing from 24 to 15 Myr) nearly offset each other's effect, and the predicted slab surface temperatures are remarkably similar throughout the subduction zone. Based on the calculated thermal

structures with phase diagrams for mafic (oceanic crust) compositions (Hacker et al., 2003b), the predicted temperature in the uppermost 500 m of the subducting oceanic crust is high enough to generate partial melting if the oceanic crust and sediments are hydrated. The thermal models predict that dehydration is nearly complete beneath the arc in southern Costa Rica, but a significant amount of water can be transported into the deep mantle below Nicaragua and NW Costa Rica, likely contributing to the along-strike variations in seismicity and arc geochemistry. However, the predicted variations in the thermal structure are insufficient to explain the large along-strike variations, leading to the conclusion that these variations are due to regional variations in sediment subduction, crustal structure, and the distribution of hydrous minerals in the incoming lithosphere (e.g., Rupke et al., 2004).

Wada and Wang (2009) calculated thermal structure of 17 subduction zones using a kinematic–dynamic model with a 100 m thick interface layer (**Figure 16(d)**). In their model, the slab–mantle interface is weakened and decoupled to 75 km depth, and downdip of this depth, the interface layer has the same rheology as the overlying mantle. In the calculations of frictional heating, σ_n is approximated by the weight of the overriding rock column and a value of 0.03 for μ' is assumed (see **Section 7.09.3.3**). In the model, the frictionally coupled part of the subduction interface where Q_{FH} is calculated (**Figure 18**) is assumed to extend roughly to the estimated maximum depth of the seismogenic zone. Below this depth, shear stress and strain rate are calculated using the viscosity and strain rate in the interface layer. They applied a dislocation-creep rheology to the interface layer over a short distance immediately downdip of the Q_{FH} segment and a uniform viscosity to the deeper part of the interface layer. Flow-law parameters for the layer dislocation-creep rheology are chosen such that the transition from Q_{FH} to Q_{VD} at a prescribed depth for a given model, with a smooth transition from the shallow Q_{FH} to the thin-layer Q_{VD} as shown in **Figure 18**. The predicted slab surface temperature depends strongly on the age of the slab (**Figure 19**); at the trench, the warmest is the young Juan de Fuca slab in northern Cascadia, and the coldest is the old Pacific slab in NE Japan, where seismological studies indicate that the seismogenic zone extends to 40–50 km depths. This wide zone of frictionally coupled zone and the relatively fast subduction rate (80 mm year^{-1}) leads to relatively large frictional heating and a rapid increase in the slab surface temperature from the trench to 40 km depth, compared with other cold-slab subduction zones, such as Izu, Kermadec, or the Marianas, where the estimated seismogenic zone extends to ~20 km depth and the subduction rate is moderate (30–60 mm year^{-1} in the margin-normal direction). At 70–80 km depth, where the slab intersects with hot flowing overlying mantle, the slab surface temperature increases in all subduction zones. However, their relative warmth at greater depths is unchanged; Izu is the coolest, and Cascadia is the warmest, indicating that the thermal processes at shallow depths have critical impact on the thermal state of the deeper part of the slab.

Syracuse et al. (2010) also developed a suite of thermal models for a global study on the correlation between the subduction zone thermal structure and subduction parameters. In their calculations, they did not include the effects of frictional heating. However, they used a mantle potential temperature of 1421.5 °C that is ~126.5 °C higher than that (1295 °C) in Wada and Wang (2009). Although the effect of frictional heating is excluded, the effect of the higher mantle potential temperature in the models by Syracuse et al. (2010) leads to slightly higher slab surface temperatures for given subduction zones than that predicted by Wada and Wang (2009). Thus, although including the effect of frictional heating has important effect on the slab surface temperature, other factors, such as the assumed depth of decoupling–coupling transition and the mantle potential temperature, can also impact the slab thermal structure.

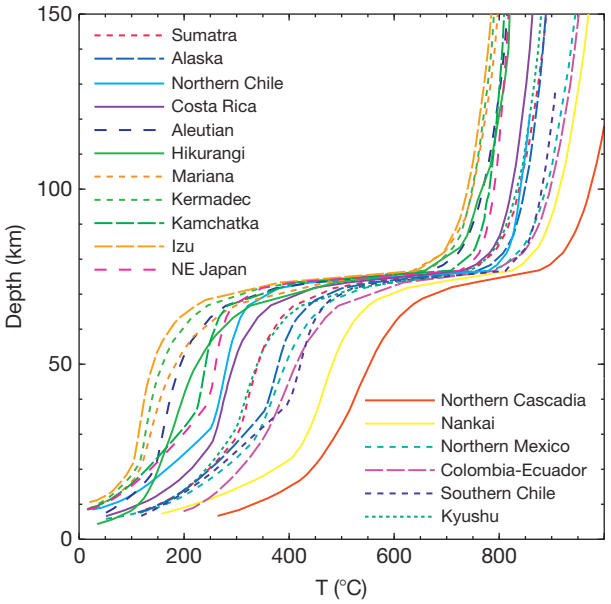


Figure 19 Temperature along the plate interface predicted by thermal models for 17 subduction zones.

7.09.4.5 Thermopetrologic Models: Fluid Availability and Subduction Zone Processes

The thermal structure of the subducting slab controls metamorphic phase changes and associated dehydration reactions, which in turn control the release of aqueous fluids at depths and volatile cycles in subduction zones. As discussed in **Sections 7.09.2.3** and **7.09.2.4**, many other processes in subduction zones, such as intraslab earthquakes, subduction thrust earthquakes, ETS, and arc volcanism, depend also on the availability of aqueous fluids. To further our understanding of the role of aqueous fluids in these processes, the pattern of fluid release from the subducting slab has been estimated based on thermopetrologic models (e.g., Hacker, 2008; Hacker et al., 2003b; Peacock, 2003; Schmidt and Poli, 2003; van Keken et al., 2011; Wada and Wang, 2009), and recent estimates indicate that 10–13 × 10^8 Tg Myr^{-1} of mineralogically bound H_2O is subducted globally and that 46–65% of the H_2O is released at depths shallower than 150 km depth (Hacker, 2008; van Keken et al., 2011) (**Figure 20**). A long-term global water cycle constrained by reconstructions of Phanerozoic sea level

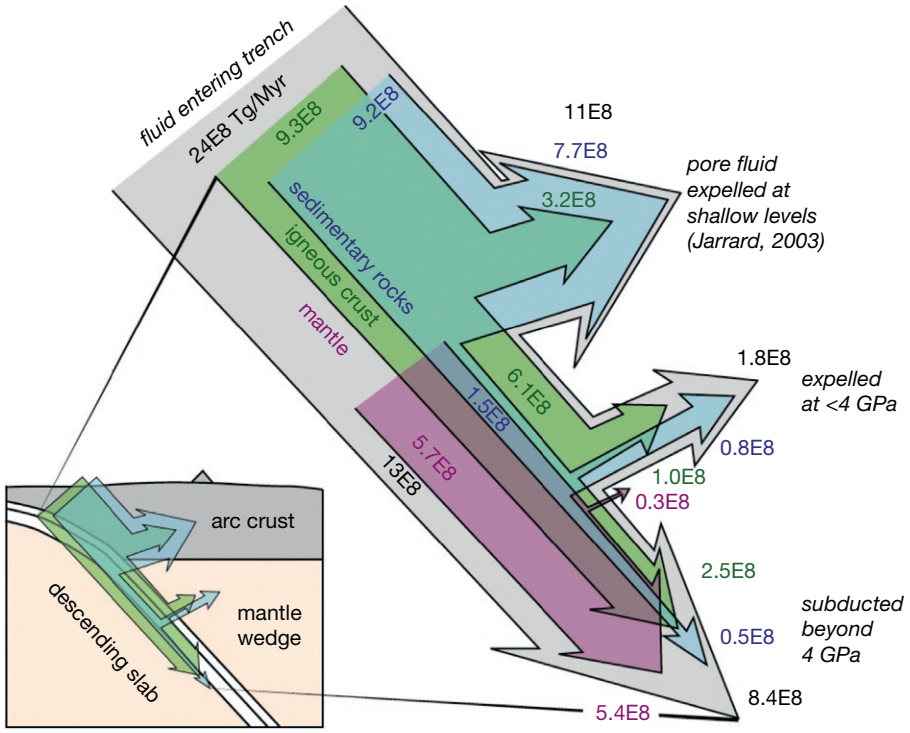


Figure 20 Global H_2O flux estimate (Hacker, 2008). The total mass of fluid entering the trench, including pore fluid, is 24×10^8 Tg Myr^{-1}. The pore fluid is expelled at shallow depths, leaving mineralogically bound H_2O of 13×10^8 Tg Myr^{-1} to be subducted. Reproduced from Hacker BR (2008) H_2O subduction beyond arcs. *Geochemistry, Geophysics, Geosystems* 9: Q03001.

change indicates that the amount of H_2O subducted beyond the subarc slab depth (>~150 km) is about half of the estimate based on the thermopetrologic models (Parai and Mukhopadhyay, 2012). It is not surprising that significant variability in the H_2O flux estimate exists, given large uncertainties in critical factors, such as the compositions and the hydration state of the incoming material, the migration path of the released H_2O, and the interaction of aqueous fluids with surrounding material.

In the fluid flux calculations based on thermopetrologic models, it is commonly assumed that the hydrated portion of the incoming lithosphere has a uniform distribution of H_2O within given lithologies. However, seismic and electrical resistivity observations (e.g., Grevemeyer et al., 2005; Ivandic et al., 2008; Key et al., 2012; Ranero et al., 2003), numerical thermomechanical and reactive flow models (e.g., Faccenda et al., 2009; Iyer et al., 2012), and exhumed sections of the oceanic crust and upper mantle (e.g., John et al., 2012; Karson, 2002) indicate that the distribution of hydrous phases in the lower crust and upper mantle is likely to be localized due to fault-controlled fluid migration and hydration. Wada et al. (2012) quantified the effect of heterogeneous hydration in the incoming plate on the pattern of fluid release using a thermopetrologic model for generic subduction zone system and found that for the same bulk H_2O content, localized hydration results in shallower H_2O release compared with uniform hydration and that the H_2O flux off the subducting slab beneath the forearc and arc regions can be almost twice as large from a locally hydrated slab as from a uniformly hydrated slab.

While subducted oceanic crust is thought to be dehydrated before reaching the transition zone (e.g., Schmidt and Poli, 1998), hydrated phases in the subducting mantle may remain stable to the transition zone along relatively cold adiabats (e.g., Staudigel and King, 1992). How much water (and other volatiles) is transported into the mantle transition zone and into the lower mantle by the subducting slab is an important question to be addressed. The estimated amount of water that is transported into the deep mantle beyond the subarc depth as mineralogically bound H_2O is about one ocean mass over the age of the Earth (van Keken et al., 2011). Nominally anhydrous olivine is an important host of hydrogen in the Earth's upper mantle and plays an important role in controlling the distribution of OH^-. The OH^- abundance in olivine also determines many important physical properties of the planet's interior, such as viscosity. The water storage capacity for olivine increases significantly with depth (Kohlstedt et al., 1996) and the transition zone phases of wadsleyite and ringwoodite have large water storage capacities (Kohlstedt et al., 1996; Smyth et al., 1997). Natural and experimentally hydrated olivine samples have been typically analyzed using uncalibrated spectroscopic methods with large uncertainties in accuracy. Bell et al. (2003) determined the hydrogen content of natural olivine samples using a new technique and indicated that an upward revision of some previous determinations by factors of between 2 and 4 is necessary. This could impact the calculations that have been performed to date. It is possible that if water can be carried into the transition zone, the transition zone could contain as much water as is present on the surface

(Bercovici and Karato, 2003; Smyth, 1994). This remains an outstanding question (see, for instance, Karato et al., 2006).

Kinematic–dynamic models are also used to study topics such as intraslab earthquakes (Hacker et al., 2003b; van Keken et al., 2012), fluid migration and melt generation in the mantle wedge (e.g., Arcay et al., 2005; Cagnioncle et al., 2007; Grove et al., 2006, 2009; Iwamori, 1998, 2007), seismic anisotropy in the mantle wedge (Kneller et al., 2005, 2007, 2008), mineral grain-size evolution (Wada et al., 2011), and the effect of anisotropic mantle viscosity on mantle flow (Lev and Hager, 2011). The effect of subducting slabs (e.g., 3-D slab geometry, slab rollback, and slab edge) on the mantle wedge flow pattern and the comparison of the results with the observations are discussed in **Sections 7.09.4.6, 7.09.5.4**, and **7.09.6**. The structure and the dynamics of the mantle wedge are reviewed by van Keken (2003) and Wiens et al. (2008).

7.09.4.6 Effects of 3-D Slab Geometry on Thermal Structure

The use of 2-D kinematic–dynamic models to study subduction zones is based on the premise that the along-strike component of mantle flow is small. This is a reasonable approximation when the convergence is margin-normal and there is little along-strike variation in the slab geometry. For subduction zones with oblique convergence, a 2-D model is developed for a cross section either normal to the strike of the margin or parallel to the convergence direction. In the former case, the margin-normal component of the convergence rate is applied as the subduction rate, and in the latter case, the convergence rate is applied as the subduction rate. It is more common to develop models for margin-normal cross sections. The comparison between 2-D and 3-D modeling results for oblique subduction shows that the margin-normal cross sections indeed provide a better approximation (i.e., a margin-normal 2-D model predicts a thermal structure that is more consistent with that predicted by a 3-D model than a convergence-parallel 2-D model) (Bengtson and van Keken, 2012).

As discussed in **Section 7.09.2.1**, the slab geometry can vary significantly along the margin of a subduction zone. Further, the margins themselves are often curved, resulting in an along-strike change in the dip direction. These changes in slab geometry lead to along-strike pressure gradients and mantle flow. The effects of 3-D mantle flow on the thermal structures of the mantle wedge and the subducting slab have been studied by 3-D kinematic–dynamic models (e.g., Bengtson and van Keken, 2012; Kneller and van Keken, 2008) and instantaneous models (e.g., Jadamec and Billen, 2010, 2012) (see **Section 7.09.6** for instantaneous models). Bengtson and van Keken (2012) tested the effect of 3-D slab geometry on the mantle flow pattern by comparing the results from 3-D modeling with those from 2-D modeling for cross sections of the 3-D models. They found that, in most cases, 2-D models provide reasonable approximations of the 3-D thermal structures except when the margin is strongly curved. The difference in the slab surface temperature determined by the 2-D and 3-D models for the curved subduction zone with convergence normal to the margin at the center (**Figure 21(a)**) is as much as 60 °C. When the subduction is oblique (**Figure 21(b)**), the

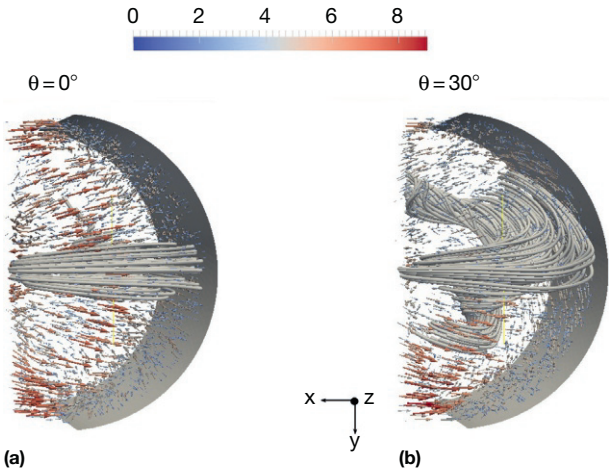


Figure 21 Mantle flow pattern (gray thick streamlines) in the mantle wedge calculated for a curved margin with (a) margin-perpendicular convergence and (b) oblique convergence (30° clockwise rotation from the margin-normal direction in (a)), at the center of the margin (Bengtson and van Keken, 2012). The color of the arrow indicates flow speed in centimeter per year. Reproduced from Bengtson AK and van Keken PE (2012) Three-dimensional thermal structure of subduction zones: Effects of obliquity and curvature. *Solid Earth* 3: 365–373.

oblique subduction gives rise to an asymmetric flow with strong toroidal component, leading to a cooler condition along the portion of the margin that has higher subduction obliquity. The 2-D model for the same region overpredicts the slab surface temperature by 120 °C. These results indicate that the interpretation of 2-D modeling results with strong 3-D slab structure needs to be interpreted carefully and that the calculation of thermal structures of some subduction zones, such as the Marianas with a strongly curved margin with oblique subduction, requires 3-D modeling. 3-D kinematic–dynamic models have also been used to study the dynamics of small-scale convection and in the mantle wedge (e.g., Honda et al., 2010; Honda and Saito, 2003; Honda and Yoshida, 2005; Zhu et al., 2011, 2009).

Although coupled dynamic–kinematic models provide description of the thermal field in the subducting slab in both 2-D and 3-D, they cannot address the dynamics of subducting slabs because the motion of the slab is prescribed. The deformation that occurs as a plate bends, breaks, and generally deforms as it passes through a subduction zone (e.g., Buffett and Rowley, 2006; Conrad and Hager, 1999) is not accounted for in kinematic models, except in so far as the model prescribes deformation by imposing a curved slab or variation in velocity. With the clear recognition of changes in slab dip, thickening of the slab, possible tears in the slab, and along slab flow, regional-scale models with the resolution of kinematic models and the ability for the slab to deform without a priori specification are needed. Further, because of the limited domain of the coupled kinematic–dynamic models (both horizontally and vertically), it is not possible to calculate geoid and dynamic topography anomalies. These are some of the motivations of applying dynamic models (e.g., Billen and Gurnis, 2001; Billen et al., 2003), as will be discussed in **Section 7.09.5**.

7.09.5 Dynamic Subduction Models

Dynamic subduction models are generally an attempt to produce a subduction-like geometry within a mantle convection calculation without prescribing plate and slab velocities, to simulate the evolution of subducting slabs. In dynamic subduction models, the plate velocity, slab velocity, and slab dip are generally not prescribed, making it more challenging to set up a calculation with a geometry that resembles a specific subduction zone. In some models, plate velocity is imposed (e.g., Billen and Hirth, 2007), and the approach is no longer fully dynamic and is referred to as a 'dynamic model with kinematic boundary conditions' (Billen, 2008). The purpose of this approach is often to investigate kinematic effects on slab evolution, though caution must be taken when interpreting the results as there may be some inconsistency between the buoyancy of the slab and the imposed surface velocity. In this section, we briefly review the simplifying approximations that are commonly made in mantle convection calculations and how plate subduction emerges in mantle convection models, and we then discuss the effects of the mantle and slab viscosities, slab rollback, trench motion, and phase transformations on the geometric evolution of the slab and the flow of the mantle that surrounds the slab. Viscoelastic or viscoelastic–plastic rheology has been applied in the dynamic models, for example, to investigate its role in controlling plate generation and slab bending. Here, we focus on slab bending and its implications to the dynamics of subduction and plate tectonics. Lastly, we review how geoid calculations are performed using dynamic models and how they compare with the observations. With the advances in numerical methods and computational capabilities, both regional- and global-scale 3-D dynamic models are now routinely calculated. In the discussions to follow, we have tried to include both 2-D and 3-D modeling results on each of the aforementioned topics.

7.09.5.1 Simplifying Approximations from Compressible to Incompressible

At depth, the mantle undergoes compression due to lithostatic pressure, and convection of the mantle results in adiabatic heating and cooling. The mechanical energy of the convecting mantle is converted into heat by viscous energy dissipation. In this section, we briefly describe the differences among the simplifying approximations that are commonly made in convection calculations. These differences are also discussed in detail by a number of authors (e.g., Ismail-Zadeh and Tackley, 2010; King et al., 2010; Chapter 7.02). Following King et al. (2010), the general forms of the equations for the mass, momentum, and energy conservation for a compressible fluid are given by

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho \, \vec{u} \right) = 0 \qquad [31]$$

$$\rho \frac{\mathrm{D} \, \vec{u}}{\mathrm{D}t} = -\nabla P + \nabla \cdot \tau + \rho \, \vec{g} \qquad [32]$$

and

$$\rho c_\mathrm{p} \frac{\mathrm{D}T}{\mathrm{D}t} - \alpha T \frac{\mathrm{D}P}{\mathrm{D}t} = \nabla \cdot (k \nabla T) + \rho H + \Phi \qquad [33]$$

respectively, where D/Dt is the material derivative, H is the volumetric heat production, and Φ is the viscous energy dissipation (see also Section 7.09.3.3 for viscous energy dissipation) (e.g., Bercovici et al., 1992, Ita and King, 1994; Jarvis and Mckenzie, 1980; King et al., 2010; Leng and Zhong, 2008; Schubert et al., 2001; Turcotte et al., 1974) (see Chapter 7.02). The deviatoric stress tensor τ is given by

$$\tau = 2\eta \dot{\varepsilon} = \eta \left(\nabla \vec{u} + \nabla \vec{u}^{T} \right) - \frac{2}{3} \eta \nabla \cdot \vec{u} \, \delta_{ij} \qquad [34]$$

and Φ is given by

$$\Phi = \tau_{ij} \frac{\partial u_i}{\partial x_j} \qquad [35]$$

In the compressible calculations, we express temperature, pressure, and density as

$$T = \overline{T} + T' \qquad [36]$$

$$P = \overline{P} + P' \qquad [37]$$

$$\rho = \overline{\rho} + \rho' \qquad [38]$$

where the overbar quantities represent time-independent reference profiles (depth-dependent) and the prime quantities represent time-dependent quantities. Assuming that $P' \ll \overline{P}$, the energy equation (eqn [33]) becomes

$$\rho c_\mathrm{p} \frac{\mathrm{D}T'}{\mathrm{D}t} = \nabla \cdot \left(k \nabla \left(\overline{T} + T' \right) \right) + \rho H + \phi - \rho c_\mathrm{p} \, \vec{u} \cdot \nabla \overline{T} - \alpha \left(\overline{T} + T' \right) \overline{\rho} g w \qquad [39]$$

where $-gw = \vec{u} \cdot \vec{g}$. In most cases, ρ' in eqn [37] is dropped, and ρ is replaced by $\overline{\rho}$ in eqns [31], [32], and [39].

The equations are nondimensionalized using reference values for temperature (T_r), distance (L), viscosity (η_r), density (ρ_r), heat capacity ($c_{\mathrm{p}_\mathrm{r}}$), and conductivity ($k_\mathrm{r}$), assuming that $c_{\mathrm{p}_\mathrm{r}} = c_\mathrm{p}$. The reference values for velocity (u_r), pressure (P_r), and time (t_r) are given by

$$u_\mathrm{r} = \frac{k_\mathrm{r}}{\rho_\mathrm{r} c_{\mathrm{p}_\mathrm{r}} L} \qquad [40]$$

$$P_\mathrm{r} = \frac{\eta_\mathrm{r} k_\mathrm{r}}{\rho_\mathrm{r} c_{\mathrm{p}_\mathrm{r}} L^2} \qquad [41]$$

$$t_\gamma = \frac{\rho_\gamma c_{\mathrm{p}_\gamma} L^2}{k_\gamma} \qquad [42]$$

Hereafter, the parameters in the equations are in their nondimensionalized forms.

Assuming an infinite Prandtl number (Pr), that the relative volume change due to temperature is $\alpha_\mathrm{r} \Delta T_\mathrm{r} \ll 1$, and that $M^2 Pr \ll 1$ (Bercovici et al., 1992), where M is the Mach number, we arrive at the anelastic liquid approximation (ALA) in which eqns [31], [32], and [39] take the forms of

$$\nabla \cdot \left(\rho \, \vec{u} \right) = 0 \qquad [43]$$

$$-\nabla P' + \nabla \cdot \tau = Ra \rho \alpha g \frac{T'}{\Delta T_\mathrm{r}} - Di \frac{\rho c_\mathrm{p}}{K_\mathrm{s} \gamma_\mathrm{r} c_\mathrm{v}} P' \qquad [44]$$

$$\rho c_\mathrm{p} \frac{\mathrm{D}T'}{\mathrm{D}t} + Di \rho \alpha w T' = \nabla \cdot \left(k \nabla \left(T' + \overline{T} \right) + \rho H + \phi \frac{Di}{Ra} \right. \qquad [45]$$

respectively, where Di and Ra are the dissipation number and the Rayleigh number given by

$$Di = \frac{\alpha_\mathrm{r} g_\mathrm{r} L}{c_{\mathrm{p}_\mathrm{r}}} \qquad [46]$$

$$Ra = \frac{\alpha_r \Delta T_r \rho_r^2 g_r L^3 c_{p_r}}{\eta_r k_r} \quad [47]$$

Under the truncated anelastic liquid approximation (TALA), the nonhydrostatic pressure in the buoyancy force in the momentum equation (eqn [44]) is ignored:

$$-\nabla P' + \nabla \cdot \tau = Ra \rho \alpha g T' \quad [48]$$

By further assuming that $\overline{T} = 0$, $\rho = 1$, $\alpha = 1$, $k = 1$, and $c_p = 1$, we arrive at the EBA, and the mass, momentum, and energy equations (eqns [43]–[45]) become

$$\nabla \cdot \vec{u} = 0 \quad [49]$$

$$-\nabla P' + \nabla \cdot \tau = Ra g T' \quad [50]$$

$$\frac{DT'}{Dt} + Diw T' = \nabla^2 T' + H + \phi \frac{Di}{Ra} \quad [51]$$

and the deviatoric stress tensor (eqn [34]) is given by

$$\tau = 2\eta \dot{\varepsilon} = \eta \left(\nabla \vec{u} + \nabla \vec{u}^T \right) \quad [52]$$

Under the BA, adiabatic and shear heating terms are dropped from the energy equation:

$$\frac{DT'}{Dt} = \nabla^2 T' + H \quad [53]$$

Equations [18] and [19] in **Section 7.09.4.1** and eqn [3] in **Section 7.09.3.1** are the dimensional forms of the mass, momentum, and energy equations under the BA (eqns [49], [50], and [53]), respectively.

In mantle convection calculations, the BA is commonly assumed (e.g., Billen and Hirth, 2007; Lee and King, 2011). With this approach, only the mantle potential temperature is advected, and the adiabatic temperature gradients are added a posteriori. Some studies (e.g., Tosi et al., 2010) use the EBA to incorporate the effects of adiabatic and shear heating (though thermal expansivity is often assumed to be constant), giving rise to adiabatic temperature gradients. With the compressible TALA and ALA formulations, the effects of the variation in density and other thermodynamic variables, such as thermal expansivity, can be included, by assuming that these quantities are functions of depth or by following an explicit equation of state or by using thermodynamic calculations (Ita and King, 1998; Nakagawa et al., 2009). These approximations have been used to investigate topics, such as the effects of adiabatic and shear heating in compressible convection (Zhang and Yuen, 1996), the effects of phase transformations on buoyancy and heat transfer (Ita and King, 1994, 1998), the stability of deep mantle structures (Tan and Gurnis, 2005, 2007), and the effect of nonhydrostatic pressure on buoyancy (Leng and Zhong, 2008, 2010; Nakagawa and Tackley, 2004; Tackley, 1998).

7.09.5.2 From Mantle Downwelling to Plate Subduction

In both 2-D and 3-D mantle convection models with isoviscous mantle rheology, the cold near-surface material simply sinks vertically into the deeper hotter material driven largely by the negative thermal buoyancy of the surface material, leading to double-sided and symmetric downwellings (**Figure 22**) (e.g., Bercovici et al., 1989; Bunge et al., 1996, 1997; Ratcliff et al., 1995, 1996; Tackley et al., 1993, 1994), unless some kind

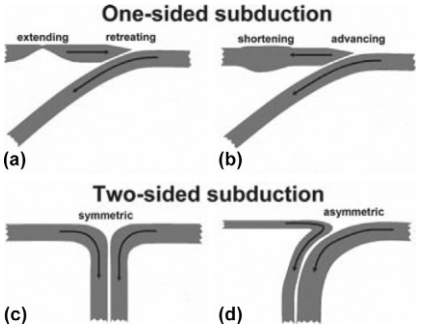


Figure 22 Schematic diagrams of (a) one-sided subduction with retreating trench, (b) one-sided subduction with advancing trench, (c) symmetric two-sided subduction, and (d) asymmetric two-sided subduction (Gerya et al., 2008). Reproduced from Gerya TV, Connolly JAD, and Yuen DA (2008) Why is terrestrial subduction one-sided? *Geology* 36: 43–46.

of asymmetric conditions are imposed. Even in those with temperature-dependent mantle rheology (some with a yield stress), the 'spontaneous' subduction is yet symmetric and double-sided (e.g., Ratcliff et al., 1995, 1996; Tackley, 1998, 2000a), unlike what we observe in subduction zones. The use of temperature-dependent rheology can also lead to stagnant-lid convection (Moresi and Solomatov, 1995; Solomatov and Moresi, 1996) (see **Section 7.09.5.3** for the effect of incorporating temperature-dependent rheology on subduction simulation).

To model the asymmetric nature of subduction in dynamic calculations, several approaches have been taken, including applications of (1) an initial condition that includes a pre-existing negatively buoyant slab at shallow depth (surface velocities may be imposed) (e.g., Christensen, 1996; Davies, 1989; Gurnis, 1992; Schmeling et al., 2008; Stevenson and Turner, 1977) (**Figure 23(a)**), (2) weak zones (e.g., Billen and Hirth, 2007; Chen and King, 1998; Gurnis, 1992; Gurnis and Hager, 1988; King and Hager, 1994; Lee and King, 2011) (**Figure 23(b)**), and (3) zero-strength faults (discontinuity) (e.g., Ita and King, 1998; Toth and Gurnis, 1998; Zhong and Gurnis, 1992, 1994, 1995) (**Figure 23(b)**); (4) density or rheological contrast between two plates with or without a vertical weak (fracture) zone separating the plates (e.g., Gerya et al., 2008; Matsumoto and Tomoda, 1983; Nikolaeva et al., 2010) (**Figure 23(c)**); and (5) self-developing initiating mechanisms, such as sediment loading and rheological weakening due to hydration (e.g., Regenauer-Lieb et al., 2001; van der Lee et al., 2008). King et al. (1992), Zhong et al. (1998), Gurnis et al. (2000b), Bercovici et al. (2000), and Bercovici (2003) discussed and compared various 'plate generation' methods (see **Chapter 7.08**). The choice of plate generation method depends on the aspect of slab evolution that is being investigated. In reality, the mechanism that leads to the observed asymmetric one-sided subduction configuration should reflect how subduction is initiated. For a detailed summary of different approaches and various mechanisms proposed for subduction initiation, refer to Gerya (2011).

The area where the asymmetry of subduction is most apparent is the shallow-dipping slabs, such as the subduction of the Pacific Plate beneath Alaska, the Juan de Fuca Plate under Washington and Oregon, and the Cocos and Nazca Plates

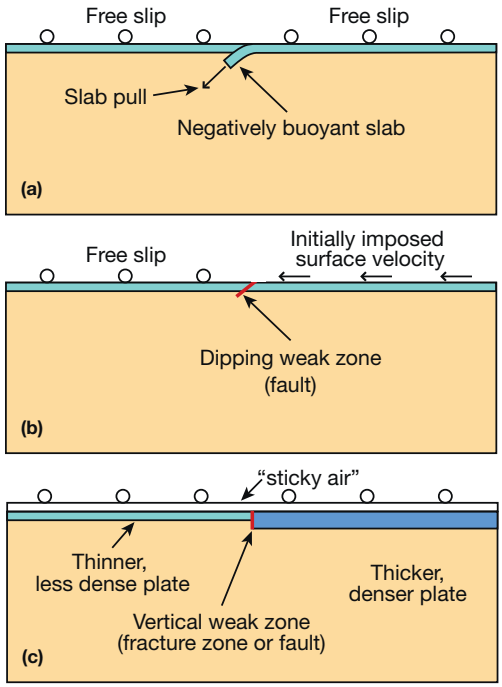


Figure 23 Examples of 2-D dynamic model setup and initial conditions with (a) a negatively buoyant slab that is attached to a surface plate, (b) a dipping weak zone representing a subduction fault, and (c) two plates of different ages juxtaposed together along a vertical weak zone, representing a fracture zone or a fault.

under Central and South America. The subduction of young lithosphere (in most cases, the slab is significantly younger than the slabs in the western Pacific) is not driven by the negative buoyancy of the slab (cf. van Hunen et al., 2000), which is the underlying assumption in most convection subduction studies. Convection models have focused on the case of older slabs where the negative buoyancy of the slab itself provides the driving motion for the plate system (e.g., slab pull). There are a few examples of studies of the subduction of young lithosphere (e.g., England and Wortel, 1980; van Hunen et al., 2000, 2001, 2002, 2004; Vlaar, 1983). Young slabs have less negative buoyancy than older slabs, and thus, younger slabs should have shallower dip angles than older slabs, all other things being equal. However, van Hunen et al. (2000, 2001) found that weak crustal rheology, a 'subduction fault' between the two plates, nonlinear rheology, and viscous shear heating all play a role in weakening the shallow crust and mantle and enabling shallow subduction. The subduction of a buoyant oceanic plateau has been suggested as a possible cause for shallow subduction; however, this does not appear to be a significant factor in the case of Peru (van Hunen et al., 2004).

In most calculations, free-slip conditions are used for the top boundary (e.g., **Figure 23(a)** and **23(b)**). With this boundary condition, the mantle is coupled with the surface, making it difficult for the cold dense surface boundary layer to detach from the surface and to sink into the hotter mantle. Imposing a free surface (e.g., Zhu et al., 2009, 2011) or a low-viscosity low-density layer, commonly called the 'sticky air' approach (e.g., Schmeling et al., 2008) (e.g., **Figure 23(c)**), to the surface boundary allows decoupling. With these approaches, plate subduction initiates without a weak zone or faults. Schmeling et al. (2008) compared these two approaches and showed that the free surface and sticky air approaches yield similar results if high resolution is used, though the latter approach results in the entrainment of the weak surface layer by subducting slab, leading to decoupling between the slab and the overlying mantle at shallow depths. Further, the sticky air approach requires careful choice of the thickness and the viscosity of the layer to simulate the effect of free surface properly (Crameri et al., 2012). It is worth pointing out that subduction formulations that explicitly include a dipping weak zone (e.g., Billen and Gurnis, 2001; Billen et al., 2003; Billen and Hirth, 2007; Chen and King, 1998; Hines and Billen, 2012; Ita and King, 1998; Lee and King, 2011; Zhong and Gurnis, 1994) do not have a problem with the cold boundary layer being unable to detach from the surface and do not need a sticky air approximation.

In 2-D dynamic calculations, for the most part, reflecting and free-slip sidewall boundary conditions are adopted. The influence of this boundary condition in 2-D is a significant limitation to subduction models because the 2-D geometry requires a strong 'return flow' that has a tendency for the slab to steepen with depth and, in many cases, dipping more than 90° – something that is not observed in subduction zones on the Earth. In 3-D Cartesian geometries, sidewalls are also a potential limitation but there is an extra degree of freedom for the flow and along-strike flow can develop (e.g., Stegman et al., 2010). Attempts to circumvent this problem have included the use of periodic boundary conditions (e.g., Gurnis and Hager, 1988; Chen and King, 1998; Gurnis and Hager, 1988; Han and Gurnis, 1999; Lowman and Jarvis, 1996) or a cylindrical geometry (Ita and King, 1998; Puster et al., 1995; Zhong and Gurnis, 1995). Symmetric, near-90°-dipping downwellings are also the observed planform in 3-D spherical convection models (e.g., Bunge et al., 1996; Tackley et al., 1994). Even when temperature-dependent rheology is included in spherical calculations, dipping slab-like features are not observed (Ratcliff et al., 1995, 1996). The use of periodic boundary conditions in two dimensions eliminates the effect of the sidewalls, and 2-D calculations can be formulated in such a way that they are formally equivalent to calculations that explicitly allow the trench to move relative to the grid (cf. Han and Gurnis, 1999). See **Section 7.09.5.4** for dynamic models that incorporate the effect of trench motion on the evolution of subducting slabs.

7.09.5.3 Effects of Phase Transformations and Thermodynamic Properties on Slab Dynamics

The fate of deep slabs beyond 660 km depth has been a topic of debate for several decades (Christensen, 1995; Lay, 1994; see **Chapter 7.12**). One of the original arguments for a barrier to convection at 660 km depth was the cessation of earthquakes at this depth (see **Chapter 4.13**). Global travel-time tomography models show fast seismic-wave speeds in the lower mantle along the Tonga subduction zone extending down to no more than 900 km in depth (Fischer et al., 1991; Grand et al., 1997; van der Hilst et al., 1997). This has led to a variety of hypotheses on what hinders the cold Tonga slab, including subhorizontal deflection of the slab (e.g., van der Hilst,

1995) or the collection of slab material at the base of the transition zone to form a large megolith (Irifune, 1993; Ringwood and Irifune, 1988). The complexity of the Tonga subduction zone and the absence of a deep fast anomaly beneath Tonga point to some form of hindrance to slabs at the base of the transition zone. It is commonly assumed that the buoyancy of a slab is dominated by the thermal buoyancy of the slab. However, it is important to recognize that phase transformations can also have significant effects locally (e.g., Bina, 1996, 1997; Christensen, 1998; Dassler and Yuen, 1996; Ita and King, 1998; Marton et al., 1999; Schmeling et al., 2008). In addition, the pattern of stresses induced by the differential buoyancy is consistent with the pattern of deep earthquakes (Bina, 1997).

The amount of slab penetration into the lower mantle provides an estimate of the rate of heat and mass transfer between the upper mantle and the lower mantle (e.g., Lay, 1994; Silver et al., 1998), and understanding slab penetration is critical to understanding thermal convection and chemical mixing of the mantle (e.g., Albarede and van der Hilst, 1999; Kellogg et al., 1999). Proposed hindrances to slab penetration into the lower mantle include the phase transformation of wadsleyite to perovskite plus magnesiowüstite (Ringwood and Irifune, 1988; Schubert and Turcotte, 1971; Schubert et al., 1975) (see **Chapter 2.08**), a change in chemical composition (e.g., Anderson, 1979; Richter, 1977), and an increase in viscosity (e.g., Hager, 1984). The presence of metastable olivine can reduce slab descent velocities by as much as 30% (Kirby et al., 1996; Marton et al., 1999; Schmeling et al., 1999), and the effect of metastable phase transformations will be greater for older and colder slabs. A compositional barrier to convection at around 660 km is not considered likely by most, and the topic of viscosity will be covered in **Section 7.09.5.4**. Here, we focus on the effect of phase transformations.

A phase change, in and of itself, will not necessarily impact convective flow. Because convection is driven by horizontal density gradients, the idealized phase transformation that happens at one pressure, P, for all temperatures will, to first order, have no impact on the flow. There is an effect due to the latent heat release due to the phase change and the small effect of dynamic pressure, which varies horizontally, but is generally on the order of 1% of the hydrostatic pressure. Phase transformations primarily affect convective flow in the mantle because of the temperature variation of the pressure of the phase transformation. This is expressed by the Clausius–Clapeyron relationship:

$$\frac{\partial P}{\partial T}=\gamma=-\frac{\Delta S}{\Delta V}=\frac{\Delta S\rho^2}{\Delta\rho}=\frac{Q_L\rho^2}{\Delta\rho T} \qquad [54]$$

where T is the absolute temperature, ΔS is the entropy change, ΔV is the specific volume change, Q_L is the latent heat, and ρ is the mean density of the two phases (see **Chapter 7.02**). A positive Clapeyron slope, such as the transformation of olivine to wadsleyite, means that the denser phase will occur at shallower depth in the slab, with the effect of increasing the negative buoyancy in the slab, enhancing convection. A negative Clapeyron slope, such as the transformation of spinel (ringwoodite) to perovskite plus ferropericlase, means that the lighter phase will be present at greater depths in the cold slab, as compared to the surrounding mantle, decreasing the net negative buoyancy of the slab and retarding convective motion (see **Figure 24**).

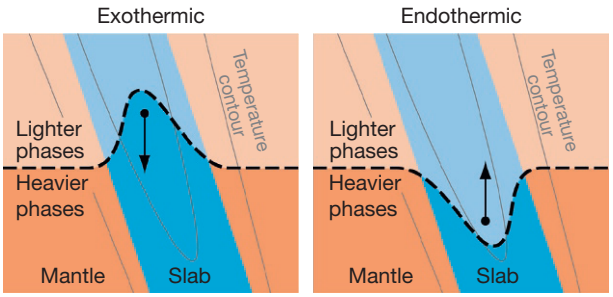


Figure 24 Cartoon of an endothermic and exothermic phase transformation within a cold slab. Modified from Figure 2 of Christensen UR (1995) Effects of phase transitions on mantle convection. *Annual Review of Earth and Planetary Sciences* 23: 65–88.

A simple comparison of densities is sufficient to illustrate the importance of phase transformations on buoyancy. The coefficient of thermal expansion for mantle minerals is on the order of $2.0\times10^{-5}\mathrm{K}^{-1}$. If we assume a 400 km column of slab material, D, that is 500 °C colder than its surrounding material, ΔT, this slab will have an integrated buoyancy anomaly, $\rho g\alpha\Delta TD$, of $132\,\mathrm{N\,m}^{-2}$. Comparing this with a phase transformation with a change in density of 6% that has a topography on the phase boundary of 30 km, the phase change topography will have a buoyancy anomaly of $0.06\rho g h_{pc}$ of $59.4\,\mathrm{N\,m}^{-2}$. Thus, a 30 km deflection of a phase boundary can have a comparable buoyancy anomaly to a 400 km column of 500 °C anomalously cold slab.

It is important to recognize that the mantle is not a single component system (**Figure 25**). Most studies of convection with phase transformations have included only the olivine system in their calculations. Transformations from pyroxene to garnet are spread out over a larger pressure range, and thus, it is generally assumed that the effect of these phase transitions will be relatively small. The dynamic effect of the olivine to wadsleyite and ringwoodite to perovskite plus ferropericlase phase changes is smaller in a pyrolite system than in a pure olivine system, because the pyroxene/garnet component makes up about 40% of the composition. Thus, the assumption of a pure olivine system in the dynamic model is likely to overestimate the effect of phase transformations on the slab dynamics.

An early attempt to model the dynamic effects of phase transformations by Christensen and Yuen (1984) employed a non-Newtonian fluid with an endothermic phase transformation. This approach produced stiff symmetric downwelling limbs that descended at a 90° dip angle as a direct consequence of the free-slip (reflecting) boundary conditions on the sides of the domain. Christensen and Yuen (1984) identified three planforms of deep slabs at a phase boundary: slab penetration, slab stagnation, and partial slab penetration. Fluid experiments with corn syrup that used a setup designed to produce an asymmetric downwelling confirm these basic planforms (Kincaid and Olson, 1987) (see also **Chapter 7.03** for the discussion of laboratory experiments). The tank experiments induce an asymmetric downwelling by placing two cold sheets of concentrated sucrose solution on the top surface of the fluid

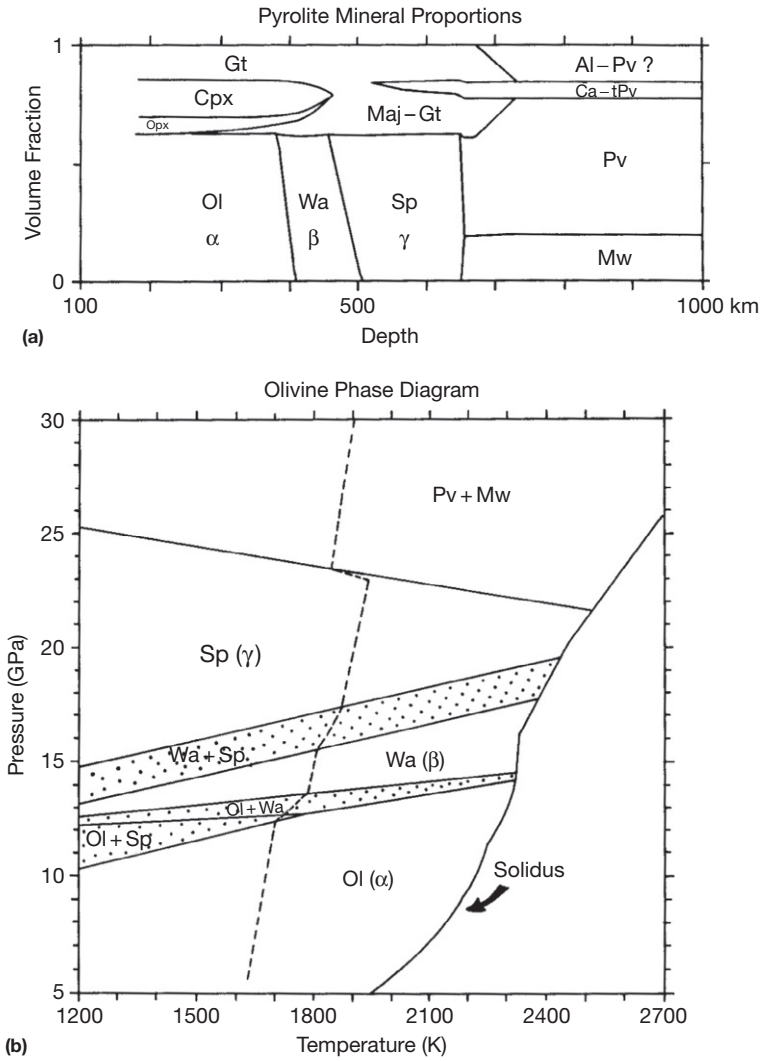


Figure 25 Phase relationship for a pyrolite upper mantle composition. (a) Volume fraction of minerals as a function of pressure. (b) Phase boundaries for the olivine component of the mantle with the dashed line as the 1573 K adiabat. Figure simplified from Ita J and Stixrude L (1992) Petrology, elasticity, and composition of the mantle transition zone. *Journal of Geophysical Research-Solid Earth* 97: 6849–6866.

in the tank. The larger of the two sheets is introduced with a shallow-dipping bend at its leading edge; this provides the instability that allows this plate to subduct under the other plate. The agreement between the tank and numerical experiments (**Figure 26**) suggests that the three planforms are robust features of cold downwellings interacting with phase transformations and/or chemical boundaries. Subsequent numerical (Ita and King, 1998; King and Ita, 1995; Zhong and Gurnis, 1994) and tank (Griffiths et al., 1995; Guillou-Frottier et al., 1995) experiments have confirmed the basic results.

Dynamic models have also demonstrated that the effect of a phase transformation on the pattern of mantle flow is sensitive to the Rayleigh number, initial conditions, boundary conditions, and the assumed equation of state (e.g., Ita and King, 1994, 1998; King et al., 1997). For example, Ita and King (1998) showed that a depth-dependent coefficient of thermal expansion approximation leads to slab stagnation and layered convection while a calculation with a self-consistent equation of state based on Ita and Stixrude (1992) and otherwise identical parameters leads to slab penetration into the lower mantle. King et al. (1997) showed that calculations including an exothermic olivine to spinel phase transformation decrease layering. Ita and King (1994) showed that at Rayleigh numbers of order 10^5, the initial flow pattern imposed on the system can persist for 30–40 transit times or more, and reflecting sidewall boundary conditions used in a Cartesian, boxlike system tends to produce layered flows more easily than those with flow-through conditions. Ita and King (1994) further showed a decrease in layering as the thermodynamic consistency is increased (i.e., Boussinesq calculations are more likely to be layered than compressible calculations with a self-consistent equation of state). The effect of latent heat on the topography of phase boundaries may also be important (cf. Christensen, 1998) although its effect is not well accounted for in most convection calculations.

Marton et al. (2005) used a thermokinetic (not dynamic) model of subduction zone thermal structure with a thermal conductivity that depends on pressure, temperature, and

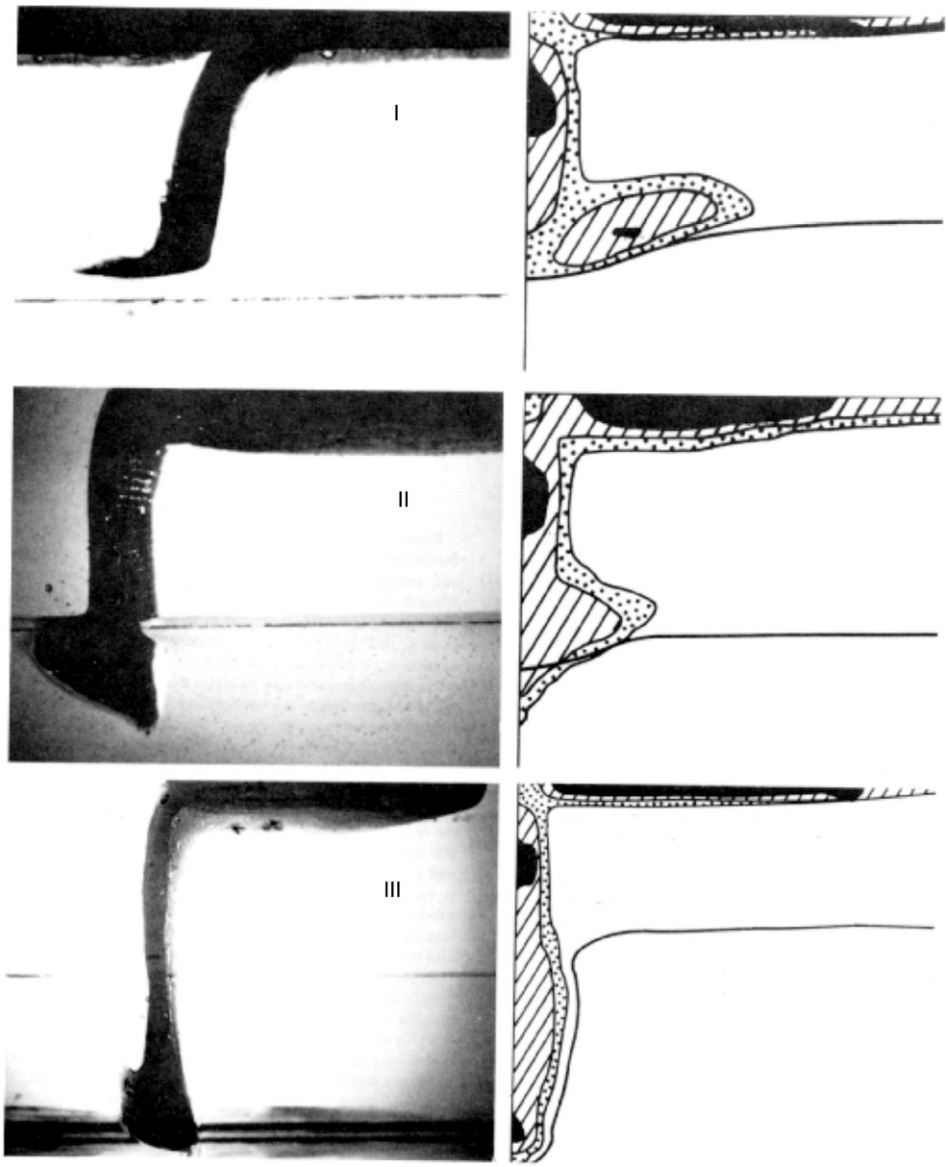


Figure 26 Experimentally observed styles of slab penetration through a density discontinuity (left) compared with two-dimensional calculation by Christensen and Yuen (1984) (right). (Part I) Slab deflection with $R \approx -0.2$. (Part II) Partial slab penetration with $R \approx 0.0$. (Part III) Complete slab penetration with $R \approx 0.5$. Reproduced from Kincaid C and Olson P (1987) An experimental-study of subduction and slab migration. *Journal of Geophysical Research-Solid Earth and Planets* 92: 13832–13840.

mineralogy. The thermal conductivity decreases exponentially with increasing temperature and the effect of pressure is much smaller, especially over the depth range of the upper mantle. However, when they add the effect of radiative heat transfer to the pressure and temperature effects of thermal conductivity, they conclude that thermal conductivity of the slab could increase by more than 50%, leading to a 50–100° increase in the temperature of the interior of the slab compared to calculations with uniform thermal conductivity. This result is independent of the slab geometry. Marton et al. (2005) used these models to address the question of olivine metastability and the ability of a metastable wedge to explain the pattern of deep seismicity. They conclude that the deepest earthquakes occur in regions that have already transformed to wadsleyite or ringwoodite, even when taking into account all of the uncertainties in their parameterization. In the case of Tonga, the deepest earthquakes extend 140 km beyond the olivine metastability region.

Nakagawa et al. (2009) incorporated the effects of thermodynamic properties, including density and thermal expansivity, on the thermochemical structure of the Earth's mantle, using a 3-D dynamic model with a diffusion-creep mantle rheology and the thermodynamic calculation code Perple_X. Although the difference between the calculated thermochemical structure and an isochemical structure that is calculated with a more conventional approach without the effects of thermodynamic properties is relatively small, the incorporation of thermodynamic properties provides a more 'realistic'

thermochemical-phase structure, from which Nakagawa et al. (2009) calculated the velocity structure of the Earth's mantle. Using a 2-D mantle convection model with the EBA formulation, Tosi et al. (2010) investigated the convection planform, focusing on the effects of increasing thermal conductivity and decreasing thermal expansivity with depth in the lower mantle based on the physical properties of postperovskite in the lower mantle. Their model assumes a Newtonian rheology and includes the effect of change in the viscosity from perovskite to postperovskite. They found that the incorporation of the effect of postperovskite leads to higher temperatures in the lower mantle and instability of the thermal boundary layer, promoting upwelling. These recent studies show the advancement in the modeling approaches to incorporate the effects of thermodynamic properties in a self-consistent manner, and similar approaches of incorporating these effects are starting to be applied to model their effects on the dynamics of subducting slabs.

7.09.5.4 Role of Mantle Rheology and Slab Strength

The rheology of the mantle is a critical factor that controls not only mantle convection but also plate tectonics and subduction dynamics. Some investigators have attempted to produce dynamic, mobile plates using complex constitutive equations (e.g., Bercovici, 1993, 1995; Stein and Hansen, 2008; Stein et al., 2004; Tackley, 1998, 2000b; Trompert and Hansen, 1998; see **Chapter 7.08**). A series of investigations of temperature-dependent Newtonian and/or non-Newtonian convection documented a transition from the free-slip planform found in constant viscosity calculations through a sluggish-lid planform to a stagnant-lid planform of convection as the stiffness of the boundary layer increases (Christensen, 1984; Moresi and Solomatov, 1995; Solomatov, 1995; Solomatov and Moresi, 1996). In stagnant- and sluggish-lid convection, most of the top boundary layer is rigid and remains at the surface and only the weakest bottom part of the boundary layer participates in the active flow. Thus, most of the cold boundary layer is never recycled into the interior of the fluid and is quite different from our picture of subduction, where most, if not all, of the subducting plate descends into the interior of the fluid. In the stagnant-lid planform, a significant amount of the negative buoyancy in the top thermal boundary layer (i.e., the plate) does not participate in the downwelling (Conrad and Hager, 1999; Moresi and Solomatov, 1995). This illustrates that there is something fundamentally unique about the dynamics of the Earth as opposed to the other terrestrial planets, where stagnant-lid convection is more applicable. While it has long been speculated that the role of water is important (e.g., Gerya et al., 2008; Lenardic and Kaula, 1993), a definitive explanation of the difference between surface deformation of the Earth and other terrestrial planets remains an open question. Furthermore, the surface heat flow from stagnant-lid convection calculations is significantly smaller than the heat flow from models with temperature-dependent rheology when a plate generation formulation (see **Section 7.09.5.2**) is included, which has a significant effect on the cooling of the Earth through time, as is commonly addressed with thermal history calculations (cf. Gurnis, 1989). The sluggish-lid and stagnant-lid convective planforms are inconsistent with piecewise-uniform surface velocities and to date, constitutive theory models have yet to produce realistic subduction zone geometries without manually initiating plate subduction. Recent studies of the role of slabs in plate kinematics are discussed in Becker and Faccenna (2009).

The addition of a strain-weakening component to the rheology weakens parts of the cold thermal boundary layer, especially at regions of high stress (such as in the corners of the computational domain). When coupled with temperature-dependent rheology (power-law rheology), this can produce nearly uniform surface velocities (King et al., 1992; van den Berg et al., 1991; Weinstein and Olson, 1992); however, power-law rheology calculations are typically unstable and the platelike behavior quickly breaks down as these calculations evolve away from the carefully chosen initial conditions. Rheologies based solely on the properties of mantle minerals require the imposition of a yield stress (Tackley, 1998, 2000a,b) or damage theory (Bercovici, 2003; Bercovici and Ricard, 2005, 2012, 2013; Bercovici et al., 2001). Some researchers argue that subduction forms at preexisting zones of weakness in the lithosphere (e.g., Kemp and Stevenson, 1996; Regenauer-Lieb and Yuen, 2000, 2003; Regenauer-Lieb et al., 2001; Schubert and Zhang, 1997; Toth and Gurnis, 1998); hence, heterogeneous properties of the lithosphere are important for dynamic subduction zone models.

The application of laboratory-derived rheological parameters for olivine (e.g., Hirth and Kohlstedt, 2003; Karato and Wu, 1993) in the dynamic slab models has become more common in the past few years. Two dominant deformation mechanisms in the upper mantle, diffusion creep and dislocation creep, follow Newtonian and power-law constitutive relations, respectively (see **Section 7.09.4.3** on mantle rheology). A composite rheological parameterization that accounts for both mechanisms is used in recent models, particularly in those that investigate the effects of the rheologies of the slab and the mantle on slab evolution (e.g., Billen and Hirth, 2007; Lee and King, 2011). It is important to recognize that at present, there are large uncertainties in the activation volume of creep laws, leading to large uncertainties in the depth dependence of viscosity due to laboratory measurements. In addition, there is very limited information about the creep properties of perovskite and magnesiowüstite. Furthermore, rheology laws are generally truncated at the highest and lowest values for computational stability. With proper care, this truncation has no impact on the solution (cf. King, 2009).

The relative strength of the mantle in the transition zone also affects the evolution of subducting slab, yet it is not well constrained. Many studies of radial viscosity structure have the lowest viscosity in the top of the upper mantle (100–400 km) while King and Masters (1992), King (1995), and others found that the transition zone is the weakest region of the upper mantle. Using 2-D Cartesian models with viscoelastic rheology that incorporates diffusion, dislocation, and the Peierls creep, Quinteros et al. (2010) studied the impact of upper mantle and transition zone viscosity on slab deformation. The Peierls creep is not commonly used in convection studies; it has the effect of reducing the stress in the cold deep slab. Their numerical approach follows that of Kameyama et al. (1999). Their modeling results show that with a weak transition zone, the slab buckles and stalls at the base of the transition zone, while

with a strong transition zone, the slab descends slowly into the lower mantle. Based on their qualitative assessment of slab geometry, they prefer a radial viscosity model with the weakest layer being the upper mantle, a slightly stronger transition zone, and a step of five to ten to the lower mantle. While they evaluated the magnitude of the Clapeyron slope between ringwoodite and perovskite plus magnesiowüstite and found this had little impact on their results, they did not systematically evaluate the effect of slab buoyancy or rheology.

The strength of subducting slabs relative to that of the surrounding mantle plays a critical role in the evolution of subducting slabs. On one hand, some of the numerical dynamic slab modeling studies constrained by slab geometry predict that relatively strong slabs that are 250–500 times stiffer than the surrounding upper mantle lead to trench retreat and back-arc–arc extension (e.g., Billen and Hirth, 2007; Di Giuseppe et al., 2008). On the other hand, some convection studies predict that slabs are weak. Houseman and Gubbins (1997) used a dynamic model of the lithosphere that assumes that the properties of the slab are uniform throughout the entirety of the slab; the slab is both more viscous and denser than the surrounding mantle. They found that the shape of the deformed slab is a strong function of the viscosity of the slab. They also observed a buckling mode that produces slab geometries similar to the Tonga slab. The effective viscosity of the slab needed to produce this slab geometry is 2.5×10^{22} Pa s, no more than 200 times the upper mantle viscosity and maybe only a factor of two greater than the lower mantle viscosity (cf. King, 1995). Using stress orientations inferred from CMT as constraints (**Section 7.09.2.3**), Alpert et al. (2010) investigated the viscosities of the slab, the upper mantle, and the lower mantle using an instantaneous flow model. Their model predicts that the slab and the lower mantle viscosities are 10–100 and 30–100 times the upper mantle viscosity, respectively. Instantaneous models are discussed in **Section 7.09.6**. Laboratory experiments show that the slab needs to be 150–500 times more viscous than the surrounding mantle to satisfy the observed slab shape (Funiciello et al., 2008). Thus, the estimates of slab strength range widely, and whether the slab is weak or strong remains an outstanding question. The role of slab strength in trench kinematics and viscous bending of the slab is discussed in **Sections 7.09.5.5** and **7.09.5.6**, respectively.

The regional gravitational potential, or geoid, has been used as a constraint on the strength of the slab (Hager, 1984). There were a number of efforts to improve the uniform viscosity flow model used by Hager by including both depth-dependent viscosity and temperature-dependent viscosity, the effect of phase transformations, and plate formulations with temperature-dependent rheology (Chen and King, 1998; King and Hager, 1994; Moresi and Gurnis, 1996; Richards and Hager, 1989; Ritzert and Jacoby, 1992; Zhong and Davies, 1999; Zhong and Gurnis, 1992). The surprising result from this work was that weak slabs provide a much better fit to the geoid than strong (high-viscosity) slabs. Additionally, King (2002) showed that dynamic support of the slab from an endothermic phase transformation is not enough to explain the geoid highs over subduction zones. The most recent results, however, illustrate that the geoid can be matched with either strong or weak slabs, ruling out the geoid as a potential constraint on slab rheology (Hines and Billen, 2012; Tosi et al., 2010). Using slab deformation as a constraint on rheology appears to hold some promise (e.g., Lee and King, 2011; Ribe et al., 2007; Stegman et al., 2010). The newest generation of instantaneous flow models (e.g., Alisic et al., 2010) could provide additional constraints on slab strength by comparing the stress states with those from seismology (see **Section 7.09.6** on instantaneous models).

Using a 3-D dynamic model, Stegman et al. (2010) investigated the effect of slab strength and buoyancy on the style of subduction. In their model, the slab is represented by a viscoplastic plate with a strong viscous core, and its strength determines the flexural stiffness of the slab. For slab buoyancy, they used the Stokes buoyancy defined as the ratio of volumetric potential energy to the upper mantle viscosity and determined the bending moment that acts on the slab. The modeling results show five different regimes of subduction depending on the slab buoyancy and effective flexural stiffness: (I) a 2-D trench retreat mode accompanied by slab rollback resulting in a flat-lying slab, (II) an advance-fold-retreat mode resulting in a flat-lying slab with a single, long recumbent fold, (III) an trench advance mode accompanied by forward plate motion resulting in an overturned, flat-lying slab, (IV) a 3-D trench retreat mode accompanied by slab rollback resulting in a flat-lying slab, and (V) a nearly stationary trench mode accompanied by a forward moving plate resulting in a slab pile (**Figure 27**). They found that 'weak' slabs, which are defined as those that are 80–100 km thick and have 100–300 times the viscosity of the surrounding mantle, exhibit modes IV and V. Based on the lack of trench advances on Earth and seismic images that indicate folding of slabs, they proposed that slabs are generally weak and are in the folding mode (Regime V). They further pointed out that if slabs are indeed weak, 3-D mantle flow can have a large effect on slab dynamics.

As subduction continues into the lower mantle, the geometry of the slab depends strongly on the relative strengths of the slab and the mantle. Ribe (2003) developed a new viscous thin-sheet theory from which he deduced scaling relationships for buckling of cold viscous material impinging on an impermeable boundary. The style of the resulting slab buckling (viscous or gravitational buckling modes) is determined by the dimensionless parameter (Ribe, 2003; Ribe et al., 2007):

$$\Pi = H_{\mathrm{d}} \left(\frac{g \Delta \overline{\rho}}{\overline{\eta} U_1 d_1^2} \right)^{1/4} \qquad [55]$$

where H_{d} is the effective fall height, g is the effective gravity, $\Delta\overline{\rho}$ is the mean density contrast between the subducting slab and ambient upper mantle, $\overline{\eta}$ is the mean slab viscosity, and U_1 and d_1 are the velocity and thickness of the buckled slab, respectively. Ribe et al. (2007) showed that scaling laws developed based on the analysis of slabs in laboratory are consistent with the thickening seen in tomographic models. The buckling amplitude and period have also been shown to be consistent with numerical experiments (e.g., Behounkova and Cizkova, 2008; Lee and King, 2011). For details, readers are advised to refer to Ribe (2003). See also **Chapter 7.04**.

7.09.5.5 Slab Rollback and Shallow-Dipping Slab

In both numerical and laboratory experiments, young buoyancy-driven slabs steepen with time from the initiation

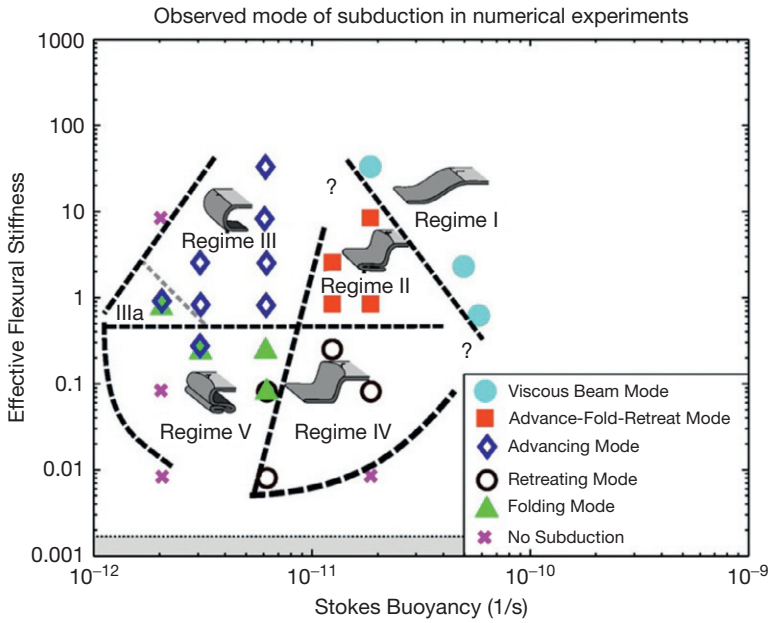


Figure 27 Five different subduction regimes predicted by 3-D modeling results of Stegman et al. (2010). The regime diagram is based on 25 numerical experiments, and the regime boundary lines are approximate. Reproduced from Stegman DR, Farrington R, Capitanio FA, and Schellart WP (2010) A regime diagram for subduction styles from 3-D numerical models of free subduction. *Tectonophysics* 483: 29–45.

of subduction until the slab reaches the transition zone (Becker et al., 1999; Griffiths et al., 1995; Gurnis and Hager, 1988; Lee and King, 2011). Once slabs reach the top of the lower mantle, slab dip angles in the upper mantle become progressively shallower if there is oceanward trench migration because the deep slab becomes anchored in the higher-viscosity lower mantle. The numerical and laboratory experiments observe slab rollback (trench retreat) while compilations of slab geometry show that older slabs are advancing, not retreating (e.g., Heuret and Lallemand, 2005).

Trench retreat (slab rollback) has been one of the most studied mechanisms for producing flat-lying slabs (Christensen, 1996; Griffiths et al., 1995; Gurnis and Hager, 1988; Ita and King, 1998; Kameyama and Nishioka, 2012; Nakakuki et al., 2010; Torii and Yoshioka, 2007; Yoshioka and Sanshadokoro, 2002; Zhong and Gurnis, 1995). Using a 2-D dynamic model with imposed slab motion and trench retreat, Kameyama and Nishioka (2012) simulated the stagnation of the slab in the transition zone and investigated the mantle flow pattern in the resulting 'Big Mantle Wedge.' Their model uses the EBA, incorporates the effects of phase transformations at 410 and 660 km depth on the density structure, and assumes an increase in viscosity by a factor of up to 30 from the upper mantle to the lower mantle. They found that a viscosity increase by a factor of 30 leads to slab stagnation at the 660 km depth regardless of the rate of trench retreat (**Figure 28**). However, without trench retreat, the stagnation leads to a simple pileup of slab material around the 660 km depth. With increasing rate of trench retreat, the horizontal extent of the flat-lying slab increases. A sufficiently fast rate of trench retreat ($>\sim$3 cm year^{-1}) results in a flat-lying slab even without an increase in mantle viscosity at the 660 km depth. They found that upwelling of the mantle develops at the hinge of the flat-lying slab (i.e., where the slab starts to become flat around the 660 km discontinuity) and proposed that this upwelling of the mantle together with the release of water from the flat-lying slab may explain the intracontinental volcanism observed in Northeast Asia where the subducted Pacific slab lies flat around the 660 km discontinuity (see **Section 7.09.2.1**). In the model of Nakakuki et al. (2010), the motion of the subducting slab, including slab rollback, is calculated within the model. This model was used to investigate the formation of flat-lying slab and its eventual descent into the lower mantle. The model incorporates the history-dependent strength of the slab, the effect of the 410 and 660 km phase transformations, using the EBA. The results show that a flat-lying slab occurs when there is sufficient 'phase buoyancy' as the trench retreats. When the slab in the transition zone is relatively weak (e.g., due to slow grain-size growth), the continued subduction causes horizontal buckling of the slab at the hinge and eventual sinking of the buckled part of the slab. As shown by these studies, numerical models generally predict a flat-lying slab when the effects of trench retreat and a viscosity increase and/or a phase transformation at the 660 km discontinuity are included.

It is interesting to note that while numerical calculations and tank experiments have demonstrated that trench retreat is an effective mechanism for generating shallow-dipping slabs in models, compilations of data from various subduction zones (Jarrard, 1986) show almost no correlation between the dip of a slab in the 100–400 km depth range and the rate of trench migration (**Figure 29**). The data from the recent compilations (e.g., Cruciani et al., 2005; Heuret and Lallemand, 2005; Lallemand et al., 2005) also indicate the lack of correlation between the slab dip and the trench migration rate, which is yet to be addressed.

A sequence of recent papers on 'slab width,' which refers to the along-strike length of the subduction zone, represent a turning point in geodynamic modeling of subduction zones that has had an important impact on the field (Schellart et al.,

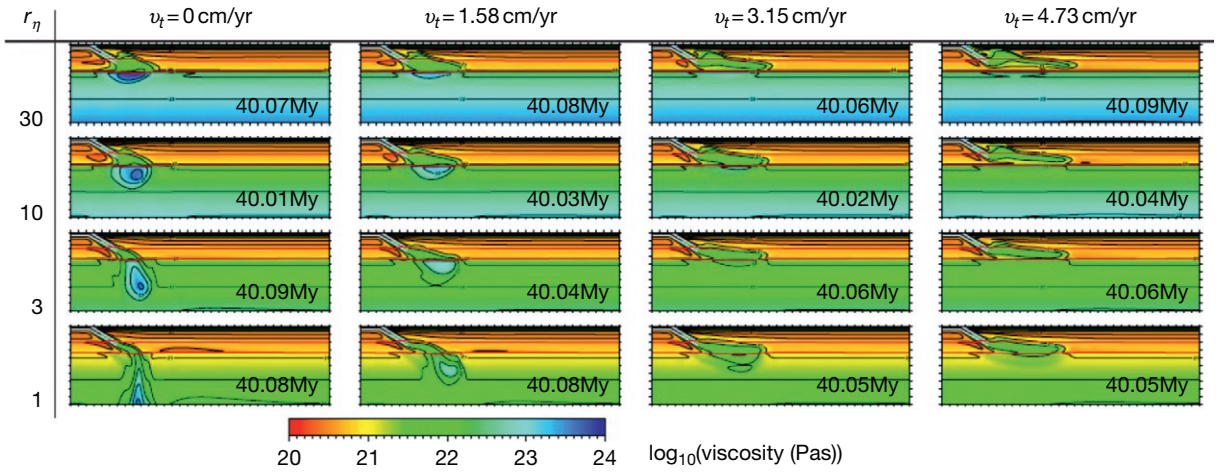


Figure 28 Modeling results from Kameyama and Nishioka (2012), showing the variability in the behavior of the subducting slab with the viscosity increase (r_η) at the 660 km discontinuity and the rate of trench retreat (v_t). Black lines are viscosity contours, and red lines represent phase boundaries. Tick marks on the horizontal and vertical axes indicate 200 km intervals. Reproduced from Kameyama M and Nishioka R (2012) Generation of ascending flows in the Big Mantle Wedge (BMW) beneath northeast Asia induced by retreat and stagnation of subducted slab. *Geophysical Research Letters* 39: L10309.

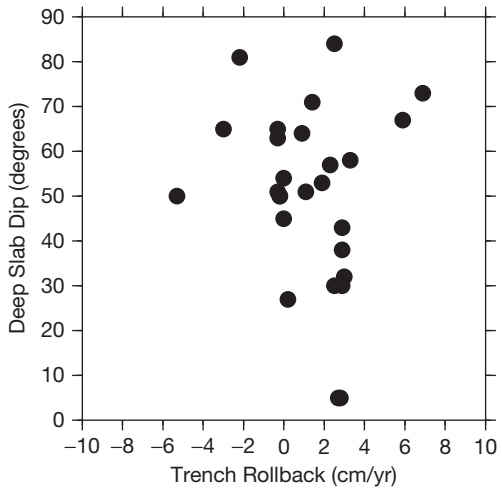


Figure 29 Deep slab dip, as measured by the shape of the Wadati–Benioff zone in the 100–400 km depth range, versus trench rollback using the Minster and Jordan plate motion model. Data taken from Jarrard (1986).

2007; Stegman et al., 2006, 2010). Ever since the first kinematic models by McKenzie and Toksöz, subduction zones and subduction zone modeling have often been visualized as 2-D structures by neglecting the along-strike dimension. However, there have been a number of studies that suggest a trench-parallel component of mantle flow above and below subducting slabs (ref) and toroidal flow around the edge of a slab (e.g., Russo and Silver, 1994; see also **Section 7.09.2.5**), and it was only recently that dynamic models were employed to conduct systematic investigations of the effect of 3-D geometry of a slab. Stegman et al. (2006) used a numerical modeling to show that as a result of the 3-D geometry of a slab, flow is allowed to accompany slab rollback around the lateral edges of the slab (the toroidal component), as opposed to 2-D geometry, where material is forced to flow underneath the slab tip (the poloidal component). This is a result that has been observed in analog tank experiments previously (e.g., Becker et al., 1999; Funiciello et al., 2003, 2004; Schellart, 2004a,b). Schellart et al. (2007) quantified two three-dimensional effects of slabs by examining 3-D numerical models: an inverse correlation between the width of the trench and the trench migration rate and the curvature of the slab. Schellart et al. (2007) showed that their results are consistent with retreat velocities observed globally, noting that maximum trench retreat velocities are observed close to slab edges and that far from edges, trench migration velocities are always slow.

7.09.5.6 Plate Bending

The discussion thus far has focused on calculations with either imposed velocities or viscous flow; however, there is another important group of calculations that include the effect of elastic strength and brittle faults (e.g., Gurnis et al., 2000b, 2004; Hall et al. 2003; Kemp and Stevenson, 1996; Regenauer-Lieb and Yuen, 2000, 2003; Regenauer-Lieb et al., 2001; Schubert and Zhang, 1997; Toth and Gurnis, 1998). The work of Vassiliou et al. (1984) and Vassiliou and Hager (1988) showed that viscous flow was able to reproduce the patterns of stress attributed to elastic bending of the plate as it entered the trench; hence, viscous flow has been thought to capture the broad pattern of deformation at a subduction zone. While this is probably true, the recognition that the deformation in the seismogenic zone (the upper 50 km or so of the boundary between the subducting plate and overriding plate) impacts the deeper slab structure by altering the slab thermal structure, and more work in this area is needed. A useful overview of the role of preexisting faults and subduction can be found in Gurnis (2000).

The bending of the plate in subduction zones is another critical process in the dynamics of plate subduction and plate tectonics. Without plate bending, subduction does not occur. Because part of the slab's GPE is partitioned into the bending

energy, slab kinematics and plate motion at the surface depend on how much bending occurs. Thicker and more viscous lithosphere is harder to bend (i.e., greater bending resistance), and with this analysis, the convective energy of the mantle dissipated by plate bending in subduction zones is a function of plate thickness and the slab's bending viscosity (e.g., Buffett, 2006; Buffett and Rowley, 2006; Conrad and Hager, 1999, 2001; Faccenna et al., 2007). Wu et al. (2008, references therein) reported the global average radius of bending curvature of 390 km based on the slab geometry delineated by intraslab seismicity. Using the radius of bending curvature, Wu et al. (2008) calculated the bending force for 207 subduction zone segments, and using an instantaneous global mantle flow model, they found that the range of bending viscosity that can explain the observed global plate motion with a reasonably small misfit (<15%) is 1–300 times the viscosity of the surrounding upper mantle, which translates into bending viscosity of 2×10^{23} Pa s or less (see **Section 7.09.6** for instantaneous models). Weaker slabs cannot transmit stresses enough to cause the observed plate motions, whereas stronger slabs have greater resistance to bending and thus plate motion. The estimated bending dissipation based on the previously mentioned bending viscosity is about 25% of the energy released by the subduction.

The dynamics of slab bending was investigated by a series of studies by Capitanio and colleagues (Capitanio and Morra, 2012; Capitanio et al., 2007, 2009). In their model, the slab is represented by a viscoelastic plate of finite width with a thin strong core, and the dynamics of the surrounding mantle is incorporated by imposing dissipative drag force and nondissipative isostatic restoring force as dynamic boundary conditions without explicitly calculating the pattern of mantle flow (Capitanio et al., 2007, references therein). The thin strong core of the slab affects the bending and stretching of the slab; a slab with the core has less resistance to bending and more resistance to stretching than an isoviscous plate of the same average viscosity. They found that the sinking velocity and the dip of the slab vary with the buoyancy and the viscosity of the slab such that the bending dissipation is minimized (Capitanio et al., 2007, 2009). Using the model, they also investigated the effects of ridge push, mantle viscosity, and the buoyancy of the overriding plate and found that the addition of slab suction and an increase in the buoyancy of the overriding plate promote trench retreat and shallower slab dip while the addition of ridge push and a reduction in mantle viscosity promote plate advance and steeper slab dip. (The sum of the plate advance and the trench retreat velocities equals the subduction velocity; Capitanio et al., 2007.) They found that the dissipation of energy due to bending is not a simple function of slab viscosity because the system minimizes the rate of dissipation by adjusting the radius of the bending curvature and the slab dip and there is an anticorrelation between the radius and the dip. Capitanio et al. (2009) showed that a similar trend is found in the global compilation of slab geometry. From these results, they reported that the lithospheric (bending and stretching) dissipation is <20% and is relatively small (Capitanio et al., 2007, 2009). Capitanio and Morra (2012) derived scaling laws for the radius of the curvature and the slab dip, using the numerical modeling results that include the effects of subduction force and far-field stresses (**Figure 30**). These results show the importance of the slab buoyancy and viscosity contrast with the surrounding to the bending and subsequent geometric evolution of the slab. Further, the model in which only slab buoyancy is applied indicates that the slab viscosity is about 200 times the upper mantle viscosity and the radius and the dip are in the ranges of 100–350 km and 30–70°, respectively. The study suggests that the effects of suction force and far-field stress control are at work beyond these ranges.

7.09.6 Instantaneous Mantle Flow Models: Geoid and Plate Motion

When the change in the mechanics of subduction and mantle convection with time is not of interest, one can obtain a snapshot of a flow field by predefining temperature, density, and viscosity structures and calculating the flow field that is in

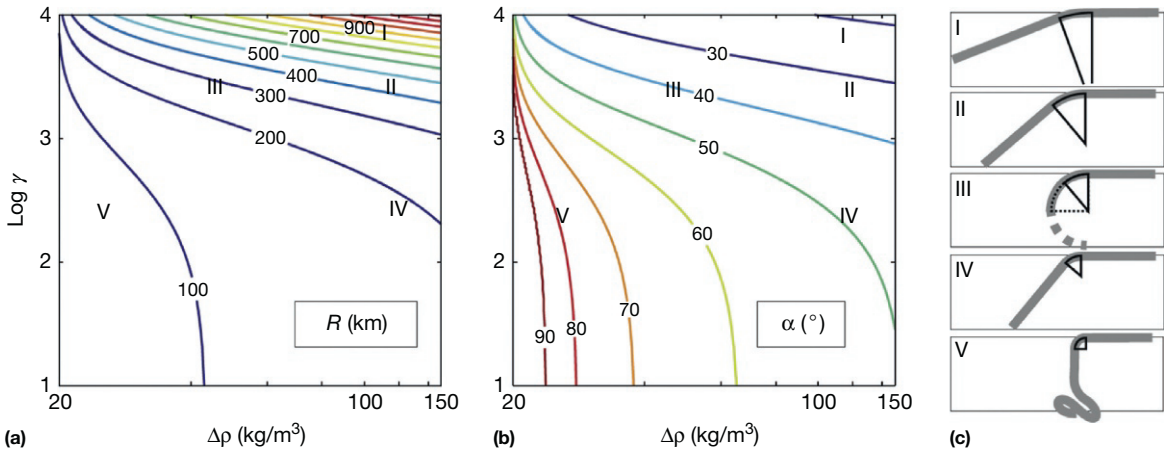


Figure 30 (a) Radius of the bending curvature and (b) slab dip as a function of density contrast between the slab and the mantle and viscosity ratio γ = [slab viscosity/mantle viscosity], and (c) different styles of subduction defined by the buoyancy and viscosity contrast: I, shallow dip (20–30) and a large radius; II, intermediate dip (30–40) and an intermediate radius (400–500 km) with trench retreat; III, slab overturn; IV, steep dip (>50) and a small radius; and V, quasi-stationary subduction with pileup. Reproduced from Capitanio FA and Morra G (2012) The bending mechanics in a dynamic subduction system: Constraints from numerical modelling and global compilation analysis. *Tectonophysics* 522: 224–234.

equilibrium with the predefined conditions. This approach is dynamic in the sense that the flow is driven by buoyancy, but the flow field is static and the density structures do not evolve with time. This approach is commonly referred to as an 'instantaneous' (Billen, 2008) or 'snapshot' (Gerya, 2011) model. It can be used to test how predefined structures based on observations, such as seismicity and seismic tomography, affect the deformation field. Instantaneous models have been used to investigate a wide range of topics, including global mantle flow, plate kinematics, and geoid, and have a history that goes back to the 1970s (e.g., Forte et al., 1993; Hager and O'Connell, 1981; Richards and Hager, 1984; Ricard et al., 1984). In this section, we focus on instantaneous subduction zone modeling studies for plate motions, mantle flow, and geoid at both regional and global scales.

Billen and colleagues (Billen and Gurnis, 2001; Billen et al., 2003) investigated the effect of low-viscosity mantle wedge on topography, gravity, and geoid, using a 3-D instantaneous viscous flow model with a model geometry and subduction parameters appropriate for the Tonga–Kermadec and Aleutian slabs. They found that the buoyancy of the slab predicted by a reasonable equation of state model is insufficient to cause the observed slab velocity and to match the geoid and topography profiles. This discrepancy raises important questions as to whether the buoyancy estimated from the equation of state is reasonable and how energy is dissipated with the bending of the slab (cf. Conrad and Hager, 2001). Billen and Gurnis (2001) chose to increase the buoyancy of the slab by 30% in their model; however, previous work has shown that there is a trade-off between slab buoyancy and slab strength. For example, Griffiths et al. (1995) and Houseman and Gubbins (1997) had shown that the deformation of a slab depends on the slab stiffness and slab buoyancy. These results suggest that we should look for additional observational constraints that can be used to address the slab stiffness. For example, Bevis (1986, 1988) used slab geometry and earthquake sources to estimate slab strength, and it might be worthwhile reinvestigating such an approach with additional earthquake data from the past two decades.

Software and computing resources have improved to the point where instantaneous models are now able to address specific and detailed slab geometries, such as the Alaskan–Aleutian subduction zone using composite Newtonian/non-Newtonian rheology (Jadamec and Billen, 2010, 2012). Perhaps not surprisingly, they found that incorporating realistic slab geometries into 3-D models of subduction leads to complex flow in the vicinity of the slab. A somewhat more surprising result is that flow in the mantle close to the slab can exceed eight times the surface velocity even while the surface velocity driven by the negatively buoyant slab is consistent with observed plate velocities. This fast mantle flow has not been observed in calculations with fixed (kinematic) slabs perhaps because of the constraint imposed by the kinematic velocity. Additionally, most previous 2-D dynamic subduction calculations do not use a composite or non-Newtonian rheology. Jadamec and Billen (2012) speculated that short slabs that are unsupported by the high-viscosity lower mantle and steepening with time are more likely to induce this pattern of localized rapid mantle flow.

Using a 3-D instantaneous model, Stadler et al. (2010) investigated the effect of the upper mantle rheology, the lower mantle viscosity structure, and the viscosity of weak zones between two plates. In 3-D models, uniform meshing is numerically impractical, and the dynamic nature of the model is intractable by static meshing. These shortcomings are now avoided by the use of adaptive mesh refinement (AMR) algorithms, in which fine mesh is regenerated automatically around large solution gradient (e.g., Alisic et al., 2010; Burstedde et al., 2009; Davies et al., 2007, 2011; Stadler et al., 2010). With AMR, Stadler et al. (2010) reported a 1 km resolution near the plate boundaries in their global mantle flow model. This is comparable to the grid resolution used in the corner of the mantle wedge in many 2-D subduction models (van Keken et al., 2008). Using the model, Stadler et al. (2010) found that the interaction of subducting slabs with the higher-viscosity lower mantle has a dramatic effect on plate motions and therefore that a fine balance between the lower mantle viscosity must be maintained. By adjusting the rheological parameter values within their uncertainties, Stadler et al. (2010) achieved rigid plate motions. Further, their modeling results show that the viscosity of the weak zone also has a strong impact on the large-scale kinematics and that viscous dissipation in the hinge zones of subducting slabs is 5–20% of the total dissipation within the mantle.

Alisic et al. (2010) examined the state of stress in slabs within instantaneous global flow models using ocean plate ages (Muller et al., 2008) and internal buoyancy calculated from seismic tomography (e.g., Ritsema et al., 2004) plus slabs constrained by seismicity using the RUM model (Gudmundsson and Sambridge, 1998). In their model, a composite rheology with a yield stress was used, and the yield stress and stress exponent were varied to minimize the misfit of the surface velocity field with and inferred plate velocities. The models reproduce the general trends in stresses in subduction zones, at least for the Bolivian and Tonga slabs. This approach has the potential to place further constraints on slab rheology. The descent of the Farallon slab beneath North America may have implications for intracontinental deformation in the Eastern United States (Forte et al., 2007, 2010) and Colorado Plateau (Moucha et al., 2008, 2009).

7.09.7 Concluding Remarks

Subduction zones are complex regions of the Earth with important physical and chemical processes that impact global geodynamics, geochemistry, and thermal evolution of the planet. While we have a fairly good general understanding about the processes occurring in subduction zones, many details (and a few major pieces) remain elusive. It is increasingly evident that future progress requires collaboration and communication among a variety of specialists. So perhaps topping any list of future directions or outstanding questions is the need for better communication between researchers of various backgrounds and expertise. While this has become a common statement in our scientific lexicon, when it comes to understanding subduction zones, it is very true. Mantle dynamic calculations have always pushed the envelope of high-performance computing (**Chapter 7.05**), and there is still a need for high resolution in numerical subduction models. For 3-D dynamic and instantaneous models, this has led to the

implementation of AMR techniques. For kinematic–dynamic models, refined meshes are also important, and because the model geometry is predefined, the mesh refinement can be done a priori. Addressing current and future geodynamics questions, such as those listed in the succeeding text, will require continuing advances in numerical techniques and an increase in the volume and range of observational constraints.

In conclusion, we list what we believe to be major remaining questions in the study of subduction zone dynamics:

1. How exactly does the slab thermal structure relate to the deformation and fluid availability in and around the slab? Are there observations in addition to forearc heat flow and seismic data that can be used to constrain the deformation, shear heating, and the slab–mantle decoupling–coupling transition in this region? Our understanding of the deformation and heat generation in the top 50−100 km of a subduction zone is still limited, and we feel that the need for further constraints is urgent, particularly for seismic and volcanic hazard mitigation. Based on numerical models constrained by surface heat flow observations and other geophysical observations, such as seismic attenuation, the slab–mantle decoupling–coupling transition tends to be at 70–80 km depth (e.g., Currie et al., 2004b; Furukawa, 1993; Syracuse et al., 2010; Wada and Wang, 2009). This transition is another critical factor that determines the temperature condition along the slab surface and the thermal structure of subduction zones as a whole, yet the mechanism that controls this transition depth remains unclear. The thermal condition acquired in the top 100 km is largely translated to greater depths, and its uncertainty impacts our understanding of the deeper part of the slab. Because a number of important processes, such as intraslab earthquakes, slab bending, and slab melting, depend on temperature, the thermal structure of subducting slabs needs to be constrained further for better understanding of these processes. The recent studies that focus on geochemical data as constraints for the slab surface temperature, such as the Ce/H_2O ratio (e.g., Cooper et al., 2012), are encouraging in this respect.
2. How strong are slabs and does slab deformation significantly affect the thermal structure of the slab via affecting the geometric evolution of the slab? For many problems of interest, slab rheology may not be the major controlling factor, but for understanding the response of slabs to phase transformations in the transition zone, understanding the role of slab buoyancy and plate bending in plate dynamics, and understanding whether slab components are able to separate (which could be significant in the chemical evolution of the mantle), slab rheology is almost certainly critical. Geoid observations correlate well with subduction zones and have been used to constrain dynamic models, though recent studies show that both strong and weak slabs are compatible with the geoid (cf. Hines and Billen, 2012; Tosi et al., 2009). On the other hand, the geometry and deformation of the slab are more strongly influenced by slab strength. Whether slabs bend, buckle, roll back, stagnate, or descend more or less undeformed depends strongly on both slab buoyancy and slab strength (e.g., Capitanio and Morra, 2012; Stegman et al., 2010). As the resolution in regional tomography and other seismic imaging techniques improves, we should be able to more clearly map slab deformation, which will provide a better constraint on slab strength.
3. What are the effects of thermodynamic properties on the evolution of subducting slabs? Many of the geodynamic models assume incompressible fluids with constant values for thermodynamic properties, such as density, thermal expansivity, and thermal diffusivity. The effects of phase transformations on density alone have been incorporated in a number of dynamic models using the Boussinesq or EBA (e.g., Kameyama and Nishioka, 2012; Nakakuki et al., 2010), but it is ultimately the combined effect of variations in the thermodynamic properties that is important. In the upper 200 km, variability in thermodynamic properties does not have a significant effect on slab thermal structure (Lee and King, 2009). Owing to the increased number of thermodynamic data on mantle minerals, such effect on the pattern of mantle convection has been incorporated in dynamic models (e.g., Nakagawa et al., 2009), and it is possible that such formulations will become implemented more regularly in the next generation of dynamic slab models, providing a thermodynamically self-consistent picture of the evolution of subducting slabs.
4. What is the pattern of fluid release from the subducting slab? How much water is carried into (and out of) the transition zone by slabs? The depth and rate of fluid release from the subducting slab are complex functions of slab thermal structure, pressure, composition, rate of subduction, and reaction kinetics, and addressing the pattern of fluid release is closely related to and is part of the reason why we aim to address the first and second questions discussed in the preceding text (i.e., slab thermal structure, deformation, and geometric evolution). Recent global flux estimates (e.g., Hacker, 2008; van Keken et al., 2011) indicate 54–35% of the subducted mineralogically bound H_2O is transported beyond the subarc depth (>150 km), though large uncertainties in the flux estimates still remain, owing to a range of unknowns, including the spatial distribution and local bulk content of H_2O in the incoming material, and the fact that it is also equally challenging to map out the migration path of the released fluids within the slab (e.g., Faccenda et al., 2012), which is an important part of the volatile recycling process. With increasing resolution of seismic observations, further improvements of the petrologic models and the flux estimates can be expected. The recognition that wadsleyite has a large water storage capacity has led to the interesting possibility that a large amount of water could reside in the transition zone. It is by no means clear that slabs retain any water by the time they reach the transition zone. The fate of water at the base of the transition zone is even more obscure (Bercovici and Karato, 2003; Richard and Bercovici, 2009; Richard et al., 2006). It is important to remember that this may not be a steady-state process. Water may be able to remain stable to the transition zone during relatively unusual periods of fast subduction (Staudigel and King, 1992).
5. How does subduction begin? While there have been some interesting developments in this area related to preexisting regions of weakness and/or lithospheric discontinuities,

there is a lot more to understand. It seems that the easiest way to make a new subduction zone is to assume preexisting subduction. This begs the question as to how the first subduction started. When subduction first began and what subduction looked like in the early Earth are also open questions (Halla et al., 2009; Moyen and van Hunen, 2012; van Hunen and Moyen, 2012). Transform faults may also play a role in subduction initiation (Gurnis, 2000; Hall et al., 2003; Lebrun et al., 2003; Toth and Gurnis, 1998). The role of the free surface or the 'sticky air' layer in initiating subduction in geodynamic models (e.g., Schmeling et al., 2008) suggests that this is an important factor, yet the fact that methods using weak faults or weak zones do not require a free surface merits further investigation.

Why does the Earth have plate tectonics while the other terrestrial planets do not? See also **Chapter 7.08**. While this has not really been a topic discussed in this chapter, it is one of the foremost questions related to our understanding of global planetary geodynamics. It is hard to imagine that we really understand the process of subduction if we cannot explain why the Earth is the only terrestrial planet with currently active subduction. While we have recipes for creating plate tectonics-like surface velocities in geodynamic models, our understanding of subduction remains far from complete. While the role of water has long been speculated as an important factor in subduction (cf. Hubbert and Rubey, 1959; Lenardic and Kaula, 1994), numerical models are now able to quantitatively address the role of water, through hydrated minerals and sediments, in subduction (e.g., Gerya et al., 2008). It is interesting to note that there have been a series of papers investigating the possibility of plate tectonics on extrasolar planets called super-Earths, or planets around other stars with masses one to ten times the Earth's mass (Foley et al., 2012; Korenaga, 2010; Stein et al., 2013; Tackley et al., 2013; Valencia and O'connell, 2009; Valencia et al., 2007; van Heck and Tackley, 2011). While the discussion has not reached a consensus, further exploration in this topic may help to address why the Earth is the only terrestrial planet with active subduction and to fully understand the process of subduction.

Acknowledgments

We thank Simon Peacock and Changyeol Lee for their helpful discussions and Peter van Keken and volume editor David Bercovici for their thoughtful reviews on the first and second editions of this chapter. This work was supported by NSF EAR-0408005. A number of the figures were produced with the GMT software package (Wessel and Smith, 1991).

References

Abers GA (2000) Hydrated subducted crust at 100–250 km depth. *Earth and Planetary Science Letters* 176: 323–330.

Abers GA (2005) Seismic low-velocity layer at the top of subducting slabs: Observations, predictions, and systematics. *Physics of the Earth and Planetary Interiors* 149: 7–29.

Abers GA, van Keken PE, Kneller EA, Ferris A, and Stachnik JC (2006) The thermal structure of subduction zones constrained by seismic imaging: Implications for slab dehydration and wedge flow. *Earth and Planetary Science Letters* 241: 387–397.

Albarede F and van der Hilst R (1999) New mantle convection models reconcile conflicting evidence. *Eos, Transactions of the American Geophysical Union* 45: 535–539.

Alisic L, Gurnis M, Stadler G, et al. (2010) Slab stress and strain rate as constraints on global mantle flow. *Geophysical Research Letters* 37: 22308.

Alpert LA, Becker TW, and Bailey IW (2010) Global slab deformation and centroid moment tensor constraints on viscosity. *Geochemistry, Geophysics, Geosystems* 11: Q12006.

Anderson DL (1979) Chemical stratification of the mantle. *Journal of Geophysical Research* 84: 6297–6298.

Anderson ML, Zandt G, Triep E, Fouch M, and Beck S (2004) Anisotropy and mantle flow in the Chile–Argentina subduction zone from shear wave splitting analysis. *Geophysical Research Letters* 31: L23608.

Arcay D, Tric E, and Doin MP (2005) Numerical simulations of subduction zones – Effect of slab dehydration on the mantle wedge dynamics. *Physics of the Earth and Planetary Interiors* 149: 133–153.

Asimow PD, Dixon JE, and Langmuir CH (2004) A hydrous melting and fractionation model for mid-ocean ridge basalts: Application to the Mid-Atlantic Ridge near the Azores. *Geochemistry, Geophysics, Geosystems* 5: Q01E16.

Asimow PD and Ghiorso MS (1998) Algorithmic modifications extending MELTS to calculate subsolidus phase relations. *American Mineralogist* 83: 1127–1132.

Austin NJ and Evans B (2007) Paleowattmeters: A scaling relation for dynamically recrystallized grain size. *Geology* 35: 343–346.

Baba T, Tanioka Y, Cummins PR, and Uhira K (2002) The slip distribution of the 1946 Nankai earthquake estimated from tsunami inversion using a new plate model. *Physics of the Earth and Planetary Interiors* 132: 59–73.

Bailey IW, Alpert LA, Becker TW, and Miller MS (2012) Co-seismic deformation of deep slabs based on summed CMT data. *Journal of Geophysical Research-Solid Earth* 117: B04404. http://dx.doi.org/10.1029/2011JB008943.

Batchelor GK (1967) *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.

Bebout GE, Scholl D, Kirby S, and Platt J (1996) *Subduction: Top to Bottom. Geophysical Monograph Series*, vol. 96. Washington, DC: American Geophysical Union.

Becker TW and Faccenna C (2009) A review of the role of subduction dynamics for regional and global plate motions. In: Lallemand S and Funiciello F (eds.) *Subduction Zone Geodynamics*, pp. 3–34. Berlin: Springer.

Becker TW, Faccenna C, O'Connell RJ, and Giardini D (1999) The development of slabs in the upper mantle: Insights from numerical and laboratory experiments. *Journal of Geophysical Research-Solid Earth* 104: 15207–15226.

Behounkova M and Cizkova H (2008) Long-wavelength character of subducted slabs in the lower mantle. *Earth and Planetary Science Letters* 275: 43–53.

Bell DR, Rossman GR, Maldener J, Endisch D, and Rauch F (2003) Hydroxide in olivine: A quantitative determination of the absolute amount and calibration of the IR spectrum. *Journal of Geophysical Research-Solid Earth* 108: 2105.

Bengtson AK and van Keken PE (2012) Three-dimensional thermal structure of subduction zones: Effects of obliquity and curvature. *Solid Earth* 3: 365–373.

Bercovici D (1993) A simple-model of plate generation from mantle flow. *Geophysical Journal International* 114: 635–650.

Bercovici D (1995) A source–sink model of the generation of plate-tectonics from non-Newtonian mantle flow. *Journal of Geophysical Research-Solid Earth* 100: 2013–2030.

Bercovici D (2003) The generation of plate tectonics from mantle convection. *Earth and Planetary Science Letters* 205: 107–121.

Bercovici D and Karato S (2003) Whole-mantle convection and the transition-zone water filter. *Nature* 425: 39–44.

Bercovici D and Ricard Y (2005) Tectonic plate generation and two-phase damage: Void growth versus grain size reduction. *Journal of Geophysical Research-Solid Earth* 110: B03401.

Bercovici D and Ricard Y (2012) Mechanisms for the generation of plate tectonics by two-phase grain-damage and pinning. *Physics of the Earth and Planetary Interiors* 202–203: 27–55.

Bercovici D and Ricard Y (2013) Plate generation with two-phase grain-damage and pinning: Source–sink model and toroidal flow. *Earth and Planetary Science Letters* 365: 275–288. http://dx.doi.org/10.1016/j.epsl.2013.02.002.

Bercovici D, Ricard Y, and Richards MA (2000) The relation between mantle dynamics and plate tectonics: A primer. In: Richards MA, Gordon RG, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*, pp. 5–46. Washington, DC: American Geophysical Union.

Bercovici D, Ricard Y, and Schubert G (2001) A two-phase model for compaction and damage 1. General theory. *Journal of Geophysical Research-Solid Earth* 106: 8887–8906.

Bercovici D, Schubert G, and Glatzmaier GA (1989) 3-Dimensional spherical-models of convection in the Earths mantle. *Science* 244: 950–955.

Bercovici D, Schubert G, and Glatzmaier GA (1992) Three-dimensional, infinite Prandtl number, compressible convection in a basally heated spherical shell. *Journal of Fluid Mechanics* 239: 683–719.

Bevis M (1986) The curvature of Wadati–Benioff zones and the torsional rigidity of subducting plates. *Nature* 323: 52–53.

Bevis M (1988) Seismic slip and down-dip strain rates in Wadati–Benioff zones. *Science* 240: 1317–1319.

Billen MI (2008) Modeling the dynamics of subducting slabs. *Annual Review of Earth and Planetary Sciences* 36: 325–356.

Billen MI and Gurnis M (2001) A low viscosity wedge in subduction zones. *Earth and Planetary Science Letters* 193: 227–236.

Billen MI, Gurnis M, and Simons M (2003) Multiscale dynamics of the Tonga–Kermadec subduction zone. *Geophysical Journal International* 153: 359–388.

Billen MI and Hirth G (2007) Rheologic controls on slab dynamics. *Geochemistry, Geophysics, Geosystems* 8: Q08012.

Bina CR (1996) Phase transition buoyancy contributions to stresses in subducting lithosphere. *Geophysical Research Letters* 23: 3563–3566.

Bina CR (1997) Patterns of deep seismicity reflect buoyancy stresses due to phase transitions. *Geophysical Research Letters* 24: 3301–3304.

Bock G, Schurr B, and Asch G (2000) High-resolution image of the oceanic Moho in the subducting Nazca plate from P–S converted waves. *Geophysical Research Letters* 27: 3929–3932.

Bostock MG, Hyndman RD, Rondenay S, and Peacock SM (2002) An inverted continental Moho and serpentinization of the forearc mantle. *Nature* 417: 536–538.

Bourdon E, Eissen JP, Monzier M, et al. (2002) Adakite-like lavas from Antisana volcano (Ecuador): Evidence for slab melt metasomatism beneath the Andean Northern volcanic zone. *Journal of Petrology* 43: 199–217.

Bowman JR and Ando M (1987) Shear-wave splitting in the upper-mantle wedge above the Tonga Subduction Zone. *Geophysical Journal of the Royal Astronomical Society* 88: 25–41.

Brudzinski MR and Chen WP (2003a) A petrologic anomaly accompanying outboard earthquakes beneath Fiji–Tonga: Corresponding evidence from broadband P and S waveforms. *Journal of Geophysical Research-Solid Earth* 108: 2299.

Brudzinski MR and Chen WP (2003b) Visualization of seismicity along subduction zones: Toward a physical basis. *Seismological Research Letters* 74: 731–738.

Brudzinski MR, Thurber CH, Hacker BR, and Engdahl ER (2007) Global prevalence of double Benioff zones. *Science* 316: 1472–1474.

Buffett BA (2006) Plate force due to bending at subduction zones. *Journal of Geophysical Research-Solid Earth* 111: B09405.

Buffett BA and Rowley DB (2006) Plate bending at subduction zones: Consequences for the direction of plate motions. *Earth and Planetary Science Letters* 245: 359–364.

Bunge HP and Grand SP (2000) Mesozoic plate-motion history below the northeast Pacific Ocean from seismic images of the subducted Farallon slab. *Nature* 405: 337–340.

Bunge HP, Richards MA, and Baumgardner JR (1996) Effect of depth-dependent viscosity on the platform of mantle convection. *Nature* 379: 436–438.

Bunge HP, Richards MA, and Baumgardner JR (1997) A sensitivity study of three-dimensional spherical mantle convection at 10(8) Rayleigh number: Effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *Journal of Geophysical Research-Solid Earth* 102: 11991–12007.

Burstedde C, Ghattas O, Stadler G, Tu T, and Wilcox LC (2009) Parallel scalable adjoint-based adaptive solution of variable-viscosity Stokes flow problems. *Computer Methods in Applied Mechanics and Engineering* 198: 1691–1700.

Byerlee J (1978) Friction of rocks. *Pure and Applied Geophysics* 116: 615–626.

Cagnioncle AM, Parmentier EM, and Elkins-Tanton LT (2007) Effect of solid flow above a subducting slab on water distribution and melting at convergent plate boundaries. *Journal of Geophysical Research-Solid Earth* 112: B09402.

Capitanio FA and Morra G (2012) The bending mechanics in a dynamic subduction system: Constraints from numerical modelling and global compilation analysis. *Tectonophysics* 522: 224–234.

Capitanio FA, Morra G, and Goes S (2007) Dynamic models of downgoing plate-buoyancy driven subduction: Subduction motions and energy dissipation. *Earth and Planetary Science Letters* 262: 284–297.

Capitanio FA, Morra G, and Goes S (2009) Dynamics of plate bending at the trench and slab-plate coupling. *Geochemistry, Geophysics, Geosystems* 10: Q04002. http://dx.doi.org/10.1029/2008GC002348.

Cassidy JF and Ellis RM (1993) S-wave velocity structure of the northern Cascadia subduction zone. *Journal of Geophysical Research-Solid Earth* 98: 4407–4421.

Chambat F, Ricard Y, and Valette B (2010) Flattening of the Earth: Further from hydrostaticity than previously estimated. *Geophysical Journal International* 183: 727–732.

Chase CG (1978) Plate kinematics – Americas, East-Africa, and rest of world. *Earth and Planetary Science Letters* 37: 355–368.

Chase CG (1979) Subduction, the geoid, and lower mantle convection. *Nature* 282: 464–468.

Chen PF, Bina CR, and Okal EA (2004) A global survey of stress orientations in subducting slabs as revealed by intermediate-depth earthquakes. *Geophysical Journal International* 159: 721–733.

Chen WR and Brudzinski MR (2001) Evidence for a large-scale remnant of subducted lithosphere beneath Fiji. *Science* 292: 2475–2479.

Chen WP and Brudzinski MR (2003) Seismic anisotropy in the mantle transition zone beneath Fiji–Tonga. *Geophysical Research Letters* 30: 1682.

Chen JN and King SD (1998) The influence of temperature and depth dependent viscosity on geoid and topography profiles from models of mantle convection. *Physics of the Earth and Planetary Interiors* 106: 75–92.

Chiu JM, Isacks BL, and Cardwell RK (1991) 3-D configuration of subducted lithosphere in the Western Pacific. *Geophysical Journal International* 106: 99–111.

Christensen U (1984) Convection with pressure-dependent and temperature-dependent non-Newtonian rheology. *Geophysical Journal of the Royal Astronomical Society* 77: 343–384.

Christensen UR (1995) Effects of phase transitions on mantle convection. *Annual Review of Earth and Planetary Sciences* 23: 65–88.

Christensen UR (1996) The influence of trench migration on slab penetration into the lower mantle. *Earth and Planetary Science Letters* 140: 27–39.

Christensen UR (1998) Dynamic phase boundary topography by latent heat effects. *Earth and Planetary Science Letters* 154: 295–306.

Christensen UR and Yuen DA (1984) The interaction of a subducting lithospheric slab with a chemical or phase-boundary. *Journal of Geophysical Research* 89: 4389–4402.

Collier JD, Helffrich GR, and Wood BJ (2001) Seismic discontinuities and subduction zones. *Physics of the Earth and Planetary Interiors* 127: 35–49.

Colombo D, Cimini GB, and deFranco R (1997) Three-dimensional velocity structure of the upper mantle beneath Costa Rica from a teleseismic tomography study. *Geophysical Journal International* 131: 189–208.

Conder JA (2005) A case for hot slab surface temperatures in numerical viscous flow models of subduction zones with an improved fault zone parameterization. *Physics of the Earth and Planetary Interiors* 149: 155–164.

Conder JA, Wiens DA, and Morris J (2002) On the decompression melting structure at volcanic arcs and back-arc spreading centers. *Geophysical Research Letters* 29(15): 17-1–17-4

Connolly JAD (2005) Computation of phase equilibria by linear programming: A tool for geodynamic modeling and its application to subduction zone decarbonation. *Earth and Planetary Science Letters* 236: 524–541.

Connolly JAD (2009) The geodynamic equation of state: What and how. *Geochemistry, Geophysics, Geosystems* 10: Q10014.

Connolly JAD and Petrini K (2002) An automated strategy for calculation of phase diagram sections and retrieval of rock properties as a function of physical conditions. *Journal of Metamorphic Geology* 20: 697–708.

Conrad CP and Hager BH (1999) Effects of plate bending and fault strength at subduction zones on plate dynamics. *Journal of Geophysical Research-Solid Earth* 104: 17551–17571.

Conrad CP and Hager BH (2001) Mantle convection with strong subduction zones. *Geophysical Journal International* 144: 271–288.

Contenti S, Gu YJ, Okeler A, and Sacchi MD (2012) Shear wave reflectivity imaging of the Nazca–South America subduction zone: Stagnant slab in the mantle transition zone? *Geophysical Research Letters* 39: L02310.

Cooper LB, Ruscitto DM, Plank T, et al. (2012) Global variations in H_2O/Ce: 1. Slab surface temperatures beneath volcanic arcs. *Geochemistry Geophysics Geosystems* 13: Q03024.

Crameri F, Schmeling H, Golabek GJ, et al. (2012) A comparison of numerical surface topography calculations in geodynamic modelling: An evaluation of the sticky air' method. *Geophysical Journal International* 189: 38–54.

Creager KC, Chiao LY, Winchester JP, and Engdahl ER (1995) Membrane strain rates in the subducting plate beneath South-America. *Geophysical Research Letters* 22: 2321–2324.

Crough ST and Jurdy DM (1980) Subducted lithosphere, hotspots, and the geoid. *Earth and Planetary Science Letters* 48: 15–22.

Cruciani C, Carminati E, and Doglioni C (2005) Slab dip vs. lithosphere age: No direct function. *Earth and Planetary Science Letters* 238: 298–310.

Currie CA (2004) *The Thermal Structure of Subduction Zones and Backarcs*. PhD Thesis, University of Victoria.

Currie CA, Cassidy JF, Hyndman RD, and Bostock MG (2004a) Shear wave anisotropy beneath the Cascadia subduction zone and western North American craton. *Geophysical Journal International* 157: 341–353.

Currie CA, Hyndman RD, Wang K, and Kostoglodov V (2002) Thermal models of the Mexico subduction zone: Implications for the megathrust seismogenic zone. *Journal of Geophysical Research-Solid Earth* 107: 2370.

Currie CA, Wang K, Hyndman RD, and He JH (2004b) The thermal effects of steady-state slab-driven mantle flow above a subducting plate: The Cascadia subduction zone and backarc. *Earth and Planetary Science Letters* 223: 35–48.

Dassler R and Yuen DA (1996) The metastable olivine wedge in fast subducting slabs: Constraints from thermo-kinetic coupling. *Earth and Planetary Science Letters* 137: 109–118.

Davies GF (1981) Regional compensation of subducted lithosphere – Effects on geoid, gravity and topography from a preliminary model. *Earth and Planetary Science Letters* 54: 431–441.

Davies GF (1989) Mantle convection model with a dynamic plate – Topography, heat-flow and gravity-anomalies. *Geophysical Journal International* 98: 461–464.

Davies DR, Davies JH, Hassan O, Morgan K, and Nithiarasu P (2007) Investigations into the applicability of adaptive finite element methods to two-dimensional infinite Prandtl number thermal and thermochemical convection. *Geochemistry, Geophysics, Geosystems* 8: Q05010.

Davies GF and Richards MA (1992) Mantle convection. *Journal of Geology* 100: 151–206.

Davies JH and Stevenson DJ (1992) Physical model of source region of subduction zone volcanics. *Journal of Geophysical Research-Solid Earth* 97: 2037–2070.

Davies DR, Wilson CR, and Kramer SC (2011) Fluidity: A fully unstructured anisotropic adaptive mesh computational modeling framework for geodynamics. *Geochemistry, Geophysics, Geosystems* 12: Q06001.

Davis EE, Wang KL, He JH, et al. (1997) An unequivocal case for high Nusselt number hydrothermal convection in sediment-buried igneous oceanic crust. *Earth and Planetary Science Letters* 146: 137–150.

Defant MJ and Drummond MS (1993) Mount-St-Helens – Potential example of the partial melting of the subducted lithosphere in a volcanic arc. *Geology* 21: 547–550.

DeMets C, Gordon RG, Argus DF, and Stein S (1994) Effect of recent revisions to the geomagnetic reversal time-scale on estimates of current plate motions. *Geophysical Research Letters* 21: 2191–2194.

Di Giuseppe E, van Hunen J, Funiciello F, Faccenna C, and Giardini D (2008) Slab stiffness control of trench motion: Insights from numerical models. *Geochemistry, Geophysics, Geosystems* 9: Q02014.

Dixon JE, Leist L, Langmuir C, and Schilling JG (2002) Recycled dehydrated lithosphere observed in plume-influenced mid-ocean-ridge basalt. *Nature* 420: 385–389.

Dragert H, Wang KL, and James TS (2001) A silent slip event on the deeper Cascadia subduction interface. *Science* 292: 1525–1528.

Eiler JM (2003) *Inside the Subduction Factory*. *Geophysical Monograph Series*, vol. 138. Washington, DC: American Geophysical Union.

Elliott T (2003) Tracers of the slab. In: *Inside the Subduction Factory*, pp. 23–43. Washington, DC: American Geophysical Union.

Engdahl RD, van der Hilst RD, and Berrocal J (1995) Imaging of subducted lithosphere beneath South America. *Geophysical Research Letters* 22: 2317–2320.

Engdahl ER, van der Hilst R, and Buland R (1998) Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bulletin of the Seismological Society of America* 88: 722–743.

England P, Engdahl R, and Thatcher W (2004) Systematic variation in the depths of slabs beneath arc volcanoes. *Geophysical Journal International* 156: 377–408.

England P and Wortel R (1980) Some consequences of the subduction of young slabs. *Earth and Planetary Science Letters* 47: 403–415.

Evans BW (1977) Metamorphism of alpine peridotite and serpentinite. *Annual Review of Earth and Planetary Sciences* 5: 397–447.

Evans BW (2004) The serpentinite multisystem revisited: Chrysotile is metastable. *International Geology Review* 46: 479–506.

Faccenda M, Burlini L, Gerya TV, and Mainprice D (2008) Fault-induced seismic anisotropy by hydration in subducting oceanic plates. *Nature* 455: 1097–U1098.

Faccenda M, Gerya TV, and Burlini L (2009) Deep slab hydration induced by bending-related variations in tectonic pressure. *Nature Geoscience* 2: 790–793.

Faccenda M, Gerya TV, Mancktelow NS, and Moresi L (2012) Fluid flow during slab unbending and dehydration: Implications for intermediate-depth seismicity, slab weakening and deep water recycling. *Geochemistry, Geophysics, Geosystems* 13: Q01010.

Faccenna C, Heuret A, Funiciello F, Lallemand S, and Becker TW (2007) Predicting trench and plate motion from the dynamics of a strong slab. *Earth and Planetary Science Letters* 257: 29–36.

Fischer KM, Creager KC, and Jordan TH (1991) Mapping the Tonga slab. *Journal of Geophysical Research-Solid Earth and Planets* 96: 14403–14427.

Fischer KM and Jordan TH (1991) Seismic strain rate and deep slab deformation in Tonga. *Journal of Geophysical Research-Solid Earth and Planets* 96: 14429–14444.

Fischer KM, Parmentier EM, Stine AR, and Wolf ER (2000) Modeling anisotropy and plate-driven flow in the Tonga subduction zone back arc. *Journal of Geophysical Research-Solid Earth* 105: 16181–16191.

Foley BJ, Bercovici D, and Landuyt W (2012) The conditions for plate tectonics on super-earths: Inferences from convection models with damage. *Earth and Planetary Science Letters* 331–332: 281–290.

Forsyth D and Uyeda S (1975) Relative importance of driving forces of plate motion. *Geophysical Journal of the Royal Astronomical Society* 43: 163–200.

Forte AM and Mitrovica JX (2001) Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data. *Nature* 410: 1049–1056.

Forte AM, Mitrovica JX, Moucha R, Simmons NA, and Grand SP (2007) Descent of the ancient Farallon slab drives localized mantle flow below the New Madrid seismic zone. *Geophysical Research Letters* 34: L04308. http://dx.doi.org/10.1029/2006GL027895.

Forte AM, Dziewonski AM, and Woodward RL (1993) Aspherical Structure of the Mantle, Tectonic Plate Motions, Nonhydrostatic Geoid, and Topography of the Core-Mantle Boundary. *Dynamics of Earth's Deep Interior and Earth Rotation* 135–166.

Forte AM, Moucha R, Simmons NA, Grand SP, and Mitrovica JX (2010) Deep-mantle contributions to the surface dynamics of the North America continent. *Tectonophysics* 481: 3–15.

Fouch MJ and Fischer KM (1996) Mantle anisotropy beneath northwest Pacific subduction zones. *Journal of Geophysical Research-Solid Earth* 101: 15987–16002.

Fouch MJ, Fischer KM, Parmentier EM, Wysession ME, and Clarke TJ (2000) Shear wave splitting, continental keels, and patterns of mantle flow. *Journal of Geophysical Research-Solid Earth* 105: 6255–6275.

Frohlich C (1989) The nature of deep-focus earthquakes. *Annual Review of Earth and Planetary Sciences* 17: 227–254.

Fukao Y, Hori S, and Ukawa M (1983) A seismological constraint on the depth of basalt eclogite transition in a subducting oceanic-crust. *Nature* 303: 413–415.

Fukao Y, Obayashi M, Inoue H, and Nenbai M (1992) Subducting slabs stagnant in the mantle transition zone. *Journal of Geophysical Research-Solid Earth* 97: 4809–4822.

Fukao Y, Obayashi M, Nakakuki T, and Grp DSP (2009) Stagnant slab: A review. *Annual Review of Earth and Planetary Sciences* 37: 19–46.

Fukao Y, Widiyantoro S, and Obayashi M (2001) Stagnant slabs in the upper and lower mantle transition region. *Reviews of Geophysics* 39: 291–323.

Funiciello F, Faccenna C, and Giardini D (2004) Role of lateral mantle flow in the evolution of subduction systems: Insights from laboratory experiments. *Geophysical Journal International* 157: 1393–1406.

Funiciello F, Faccenna C, Giardini D, and Regenauer-Lieb K (2003) Dynamics of retreating slabs: 2. Insights from three-dimensional laboratory experiments. *Journal of Geophysical Research-Solid Earth* 108(B4): 2207. http://dx.doi.org/10.1029/2001JB000896.

Funiciello F, Faccenna C, Heuret A, et al. (2008) Trench migration, net rotation and slab–mantle coupling. *Earth and Planetary Science Letters* 271: 233–240.

Furukawa Y (1993) Depth of the decoupling plate interface and thermal structure under arcs. *Journal of Geophysical Research-Solid Earth* 98: 20005–20013.

Gaetani GA and Grove TL (1998) The influence of water on melting of mantle peridotite. *Contributions to Mineralogy and Petrology* 131: 323–346.

Gaetani GA and Grove TL (2003) Experimental constraints on melt generation in the mantle wedge. *Inside the subduction factory*, pp. 107–134: .

Gao W, Matzel E, and Grand SP (2006) Upper mantle seismic structure beneath eastern Mexico determined from P and S waveform inversion and its implications. *Journal of Geophysical Research* 111. http://dx.doi.org/10.1029/2006JB004304.

Garfunkel Z, Anderson CA, and Schubert G (1986) Mantle circulation and the lateral migration of subducted slabs. *Journal of Geophysical Research-Solid Earth and Planets* 91: 7205–7223.

Gerya T (2011) Future directions in subduction modeling. *Journal of Geodynamics* 52: 344–378.

Gerya TV, Connolly JAD, and Yuen DA (2008) Why is terrestrial subduction one-sided? *Geology* 36: 43–46.

Gerya TV, Connolly JAD, Yuen DA, Gorczyk W, and Capel AM (2006) Seismic implications of mantle wedge plumes. *Physics of the Earth and Planetary Interiors* 156: 59–74.

Ghiorso MS and Sack RO (1995) Chemical mass-transfer in magmatic processes. 4. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid–solid equilibria in magmatic systems at elevated-

temperatures and pressures. *Contributions to Mineralogy and Petrology* 119: 197–212.
Gorczyk W, Gerya TV, Connolly JAD, and Yuen DA (2007) Growth and mixing dynamics of mantle wedge plumes. *Geology* 35: 587–590.
Gorczyk W, Gerya TV, Connolly JAD, Yuen DA, and Rudolph M (2006) Large-scale rigid-body rotation in the mantle wedge and its implications for seismic tomography. *Geochemistry, Geophysics, Geosystems* 7: Q05018.
Grand SP (2002) Mantle shear-wave tomography and the fate of subducted slabs. *Philosophical Transactions of the Royal Society of London Series A –Mathematical Physical and Engineering Sciences* 360: 2475–2491.
Grand SP, van der Hilst RD, and Widiyantoro S (1997) Global seismic tomography: A snapshot of convection in the Earth. *Geological Society of America Today* 7: 1–7.
Green DH (1973) Experimental melting studies on a model upper mantle composition at high-pressure under water-saturated and water-undersaturated conditions. *Earth and Planetary Science Letters* 19: 37–53.
Green HW and Burnley PC (1989) A new self-organizing mechanism for deep-focus earthquakes. *Nature* 341: 733–737.
Green TH and Ringwood AE (1968) Crystallization of basalt and andesite under high pressure hydrous conditions. *Earth and Planetary Science Letters* 3: 481–489.
Grevemeyer I, Kaul N, Diaz-Naveas JL, et al. (2005) Heat flow and bending-related faulting at subduction trenches: Case studies offshore of Nicaragua and Central Chile. *Earth and Planetary Science Letters* 236: 238–248.
Griffiths RW, Hackney RI, and Vanderhilst RD (1995) A laboratory investigation of effects of trench migration on the descent of subducted slabs. *Earth and Planetary Science Letters* 133: 1–17.
Gripp AE and Gordon RG (2002) Young tracks of hotspots and current plate velocities. *Geophysical Journal International* 150: 321–361.
Grove TL, Chatterjee N, Parman SW, and Medard E (2006) The influence of H_2O on mantle wedge melting. *Earth and Planetary Science Letters* 249: 74–89.
Grove TL, Till CB, Lev E, Chatterjee N, and Medard E (2009) Kinematic variables and water transport control the formation and location of arc volcanoes (vol. 459, p. 694, 2009). *Nature* 460: 1044.
Gudmundsson O and Sambridge M (1998) A regionalized upper mantle (RUM) seismic model. *Journal of Geophysical Research-Solid Earth* 103: 7121–7136.
Guillou-Frottier L, Buttles J, and Olson P (1995) Laboratory experiments on the structure of subducted lithosphere. *Earth and Planetary Science Letters* 133: 19–34.
Gurnis M (1989) A reassessment of the heat transport by variable viscosity convection with plates and lids. *Geophysical Research Letters* 16(2): 179–182.
Gurnis M (1992) Rapid continental subsidence following the initiation and evolution of subduction. *Science* 255: 1556–1558.
Gurnis M (2000) On the competing roles of fault reactivation and brittle failure in generating plate tectonics from mantle convection. In: Richards MA, Gordon RG, and van der Hilst R (eds.) *History and Dynamics of Global Plate Motions*, pp. 73–94. Washington, DC: American Geophysical Union.
Gurnis M and Hager BH (1988) Controls of the structure of subducted slabs. *Nature* 335: 317–321.
Gurnis M, Ritsema J, van Heijst HJ, and Zhong SJ (2000a) Tonga slab deformation: The influence of a lower mantle upwelling on a slab in a young subduction zone. *Geophysical Research Letters* 27: 2373–2376.
Gurnis M, Zhong SJ, and Toth J (2000b) On the competing roles of fault reactivation and brittle failure in generating plate tectonics from mantle convection. In: Richards MA, Gordon RG, and van der Hilst R (eds.) *The History and Dynamics of Global Plate Motions*, pp. 73–94. Washington, DC: American Geophysical Union.
Gurnis M, Hall C, and Lavier L (2004) Evolving force balance during incipient subduction. *Geochemistry Geophysics Geosystems* 5(7). http://dx.doi.org/10.1029/2003GC000681.
Gutenberg B and Richter CF (1939) Depth and geographical distribution of deep-focus earthquakes. *Geological Society of America Bulletin* 50: 1511–1528.
Hacker BR (2008) H_2O subduction beyond arcs. *Geochemistry, Geophysics, Geosystems* 9: Q03001.
Hacker BR, Abers GA, and Peacock SM (2003a) Subduction factory – 1. Theoretical mineralogy, densities, seismic wave speeds, and H_2O contents. *Journal of Geophysical Research-Solid Earth* 108: 2029.
Hacker BR, Peacock SM, Abers GA, and Holloway SD (2003b) Subduction factory – 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *Journal of Geophysical Research-Solid Earth* 108: 2030.
Hager BH (1984) Subducted slabs and the geoid – Constraints on mantle rheology and flow. *Journal of Geophysical Research* 89: 6003–6015.
Hager BH and Clayton RW (1989) Constraints on the structure of mantle convection using seismic observations, flow models, and the geoid. In: Peltier WR (ed.) *Mantle Convection*, pp. 658–758. Oxford: Pergamon Press.
Hager BH and O'Connell RJ (1981) A simple global-model of plate dynamics and mantle convection. *Journal of Geophysical Research* 86: 4843–4867.
Hager BH and Richards MA (1989) Long-wavelength variations in earths geoid – Physical models and dynamical implications. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 328: 309–327.
Hall CE, Fischer KM, Parmentier EM, and Blackman DK (2000) The influence of plate motions on three-dimensional back arc mantle flow and shear wave. *Journal of Geophysical Research-Solid Earth* 105: 28009–28033.
Hall CE, Gurnis M, Sdrolias M, Lavier LL, and Müller RD (2003) Catastrophic initiation of subduction following forced convergence across fracture zones. *Earth and Planetary Science Letters* 212: 15–30.
Halla J, van Hunen J, Heilimo E, and Holtta P (2009) Geochemical and numerical constraints on Neoarchean plate tectonics. *Precambrian Research* 174: 155–162.
Hamburger MW and Isacks BL (1987) Deep earthquakes in the Southwest Pacific – A tectonic interpretation. *Journal of Geophysical Research-Solid Earth and Planets* 92: 13841–13854.
Han LJ and Gurnis M (1999) How valid are dynamic models of subduction and convection when plate motions are prescribed? *Physics of the Earth and Planetary Interiors* 110: 235–246.
Healy D, Reddy SM, Timms NE, Gray EM, and Brovarone AV (2009) Trench-parallel fast axes of seismic anisotropy due to fluid-filled cracks in subducting slabs. *Earth and Planetary Science Letters* 283: 75–86.
Hebert LB, Antoshechkina P, Asimow P, and Gurnis M (2009) Emergence of a low-viscosity channel in subduction zones through the coupling of mantle flow and thermodynamics. *Earth and Planetary Science Letters* 278: 243–256.
Helffrich GR, Stein S, and Wood BJ (1989) Subduction zone thermal structure and mineralogy and their relationship to seismic-wave reflections and conversions at the slab mantle interface. *Journal of Geophysical Research-Solid Earth and Planets* 94: 753–763.
Heuret A and Lallemand S (2005) Plate motions, slab dynamics and back-arc deformation. *Physics of the Earth and Planetary Interiors* 149: 31–51.
Hilairet N, Reynard B, Wang YB, et al. (2007) High-pressure creep of serpentine, interseismic deformation, and initiation of subduction. *Science* 318: 1910–1913.
Hines JM and Billen MI (2012) Sensitivity of the short- to intermediate-wavelength geoid to rheologic structure in subduction zones. *Journal of Geophysical Research-Solid Earth* 117: B05410.
Hiramatsu Y and Ando M (1996) Seismic anisotropy near source region in subduction zones around Japan. *Physics of the Earth and Planetary Interiors* 95: 237–250.
Hirose H and Obara K (2005) Repeating short- and long-term slow slip events with deep tremor activity around the Bungo channel region, southwest Japan. *Earth, Planets and Space* 57: 961–972.
Hirose H and Obara K (2006) Short-term slow slip and correlated tremor episodes in the Tokai region, central Japan. *Geophysical Research Letters* 33: 5.
Hirschmann MM (2006) Water, melting, and the deep Earth H(2)O cycle. *Annual Review of Earth and Planetary Sciences* 34: 629–653.
Hirth G and Kohlstedt DL (1996) Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144: 93–108.
Hirth G and Kohlstedt DL (2003) Rheology of the upper mantle and mantle wedge: A view from the experimentalists. In: Eiler JM (ed.) *Inside the Subduction Factory*, pp. 83–105. Washington, DC: American Geophysical Union.
Holt WE (1995) Flow-fields within the Tonga slab determined from the moment tensors of deep earthquakes. *Geophysical Research Letters* 22: 989–992.
Honda S (1985) Thermal structure beneath Tohoku, Northeast Japan – A case-study for understanding the detailed thermal structure of the subduction zone. *Tectonophysics* 112: 69–102.
Honda S, Gerya T, and Zhu GZ (2010) A simple three-dimensional model of thermo-chemical convection in the mantle wedge. *Earth and Planetary Science Letters* 290: 311–318.
Honda S and Saito M (2003) Small-scale convection under the back-arc occurring in the low viscosity wedge. *Earth and Planetary Science Letters* 216: 703–715.
Honda S and Yoshida T (2005) Effects of oblique subduction on the 3-D pattern of small-scale convection within the mantle wedge. *Geophysical Research Letters* 32: L13307.
Hori S (1990) Seismic-waves guided by untransformed oceanic-crust subducting into the mantle – The case of the Kanto District, Central Japan. *Tectonophysics* 176: 355–376.
Hori S, Inoue H, Fukao Y, and Ukawa M (1985) Seismic detection of the untransformed basaltic oceanic-crust subducting into the mantle. *Geophysical Journal of the Royal Astronomical Society* 83: 169–197.
Houseman GA and Gubbins D (1997) Deformation of subducted oceanic lithosphere. *Geophysical Journal International* 131: 535–551.

Hubbert MK and Rubey WW (1959) Role of fluid pressure in mechanics of overthrust faulting.1. Mechanics of fluid-filled porous solids and its application to overthrust faulting. *Geological Society of America Bulletin* 70: 115–166.
Hyndman RD and Wang K (1993) Thermal constraints on the zone of major thrust earthquake failure – The Cascadia subduction zone. *Journal of Geophysical Research-Solid Earth* 98: 2039–2060.
Igarashi T, Matsuzawa T, Umino N, and Hasegawa A (2001) Spatial distribution of focal mechanisms for interplate and intraplate earthquakes associated with the subducting Pacific plate beneath the northeastern Japan arc: A triple-planed deep seismic zone. *Journal of Geophysical Research-Solid Earth* 106: 2177–2191.
Irifune T (1993) Phase transformations in the Earth's mantle and subducting slabs: Implications for their compositions, seismic velocity and density structures and dynamics. *Island Arc* 2: 55–71.
Isacks B and Molnar P (1969) Focal mechanisms of deep and intermediate-depth earthquakes and dynamics of lithosphere. *Transactions-American Geophysical Union* 50(4): 238.
Isacks B and Molnar P (1971) Distribution of stresses in descending lithosphere from a global survey of focal-mechanism solutions of mantle earthquakes. *Reviews of Geophysics and Space Physics* 9: 103–174.
Isacks B, Page R, and Lomnitz C (1968) Comment on paper by Lomnitz, C – Statistical prediction of earthquakes. *Reviews of Geophysics* 6: 99–100.
Ishikawa T and Nakamura E (1994) Origin of the slab component in arc lavas from across-arc variation of B and Pb isotopes. *Nature* 370: 205–208.
Ismail-Zadeh A and Tackley PJ (2010) *Computational Methods for Geodynamics*. Cambridge: Cambridge University Press.
Ita J and King SD (1994) Sensitivity of convection with an endothermic phase-change to the form of governing equations, initial conditions, boundary-conditions, and equation of state. *Journal of Geophysical Research-Solid Earth* 99: 15919–15938.
Ita J and King SD (1998) The influence of thermodynamic formulation on simulations of subduction zone geometry and history. *Geophysical Research Letters* 25: 1463–1466.
Ita J and Stixrude L (1992) Petrology, elasticity, and composition of the mantle transition zone. *Journal of Geophysical Research-Solid Earth* 97: 6849–6866.
Ivandic M, Grevemeyer I, Berhorst A, Flueh ER, and McIntosh K (2008) Impact of bending related faulting on the seismic properties of the incoming oceanic plate offshore of Nicaragua. *Journal of Geophysical Research-Solid Earth* 113: B05410.
Iwamori H (1998) Transportation of H_2O and melting in subduction zones. *Earth and Planetary Science Letters* 160: 65–80.
Iwamori H (2007) Transportation of H_2O beneath the Japan arcs and its implications for global water circulation. *Chemical Geology* 239: 182–198.
Iyer K, Rupke LH, Morgan JP, and Grevemeyer I (2012) Controls of faulting and reaction kinetics on serpentinization and double Benioff zones. *Geochemistry, Geophysics, Geosystems* 13: Q09010.
Jadamec MA and Billen MI (2010) Reconciling surface plate motions with rapid three-dimensional mantle flow around a slab edge. *Nature* 465: 338–U389.
Jadamec MA and Billen MI (2012) The role of rheology and slab shape on rapid mantle flow: Three-dimensional numerical models of the Alaska slab edge. *Journal of Geophysical Research-Solid Earth* 117: B02304.
Jarrard RD (1986) Relations among subduction parameters. *Reviews of Geophysics* 24: 217–284.
Jarvis GT and Mckenzie DP (1980) Convection in a compressible fluid with infinite Prandtl number. *Journal of Fluid Mechanics* 96: 515–583.
John T, Gussone N, Podladchikov YY, et al. (2012) Volcanic arcs fed by rapid pulsed fluid flow through subducting slabs. *Nature Geoscience* 5: 489–492.
Johnson MC and Plank T (1999) Dehydration and melting experiments constrain the fate of subducted sediments. *Geochemistry, Geophysics, Geosystems* 1: 1007.
Jung H (2011) Seismic anisotropy produced by serpentine in mantle wedge. *Earth and Planetary Science Letters* 307: 535–543.
Jung H and Karato S (2001) Water-induced fabric transitions in olivine. *Science* 293: 1460–1463.
Kameyama M and Nishioka R (2012) Generation of ascending flows in the Big Mantle Wedge (BMW) beneath northeast Asia induced by retreat and stagnation of subducted slab. *Geophysical Research Letters* 39: L10309.
Kameyama M, Yuen DA, and Karato SI (1999) Thermal-mechanical effects of low-temperature plasticity (the Peierls mechanism) on the deformation of a viscoelastic shear zone. *Earth and Planetary Science Letters* 168: 159–172.
Kaminski E and Ribe NM (2001) A kinematic model for recrystallization and texture development in olivine polycrystals. *Earth and Planetary Science Letters* 189: 253–267.
Kanamori H, Anderson DL, and Heaton TH (1998) Frictional melting during the rupture of the 1994 Bolivian earthquake. *Science* 279: 839–842.
Kao H, Shan SJ, Dragert H, et al. (2005) A wide depth distribution of seismic tremors along the northern Cascadia margin. *Nature* 436: 841–844.
Karason H (2002) *Constraints on Mantle Convection from Seismic Tomography and Flow Modeling*. PhD Thesis, MIT.
Karason H and van der Hilst RD (2001) Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP, P-diff). *Journal of Geophysical Research-Solid Earth* 106: 6569–6587.
Karato SI (2011) Rheological structure of the mantle of a super-Earth: Some insights from mineral physics. *Icarus* 212: 14–23.
Karato S, Jung H, Katayama I, and Skemer P (2008) Geodynamic significance of seismic anisotropy of the upper mantle: New insights from laboratory studies. *Annual Review of Earth and Planetary Sciences* 36: 59–95.
Karato S and Wu P (1993) Rheology of the upper mantle – A synthesis. *Science* 260: 771–778.
Karato SI, Bercovici D, Leahy G, Richard G, and Jing Z (2006) The transition-zone water filter model for global material circulation: where do we stand? *Earth's Deep Water Cycle* 289–313.
Karson JA (2002) Geological structure of the uppermost oceanic crust created at fast- to intermediate-rate spreading centers. *Annual Review of Earth and Planetary Sciences* 30: 347–384.
Katayama I, Hirauchi H, Michibayashi K, and Ando J (2009) Trench-parallel anisotropy produced by serpentine deformation in the hydrated mantle wedge. *Nature* 461: 1114–U1209.
Katayama I and Karato S (2006) Effect of temperature on the B- to C-type olivine fabric transition and implication for flow pattern in subduction zones. *Physics of the Earth and Planetary Interiors* 157: 33–45.
Kaula WM (1968) *An Introduction to Planetary Physics, the Terrestrial Planets*. New York: John Wiley & Son.
Kaula WM (1972) Global gravity and tectonics. In: Robertson EC (ed.) *The Nature of the Solid Earth*, pp. 386–405. New York: McGraw-Hill.
Kawakatsu H and Watada S (2007) Seismic evidence for deep-water transportation in the mantle. *Science* 316: 1468–1471.
Kelemen PB, Rilling J, Parmentier EM, Mehl L, and Hacker BH (2003) Thermal structure due to solid-state flow in the mantle wedge beneath arcs. In: Eiler JM (ed.) *Inside the Subduction Factory*, pp. 293–311. Washington, DC: American Geophysical Union.
Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283: 1881–1884.
Kemp DV and Stevenson DJ (1996) A tensile, flexural model for the initiation of subduction. *Geophysical Journal International* 125: 73–94.
Key K, Constable S, Matsuno T, Evans RL, and Myer D (2012) Electromagnetic detection of plate hydration due to bending faults at the Middle America Trench. *Earth and Planetary Science Letters* 351: 45–53.
Kincaid C and Olson P (1987) An experimental-study of subduction and slab migration. *Journal of Geophysical Research-Solid Earth and Planets* 92: 13832–13840.
King SD (1995) Radial models of mantle viscosity – Results from a genetic algorithm. *Geophysical Journal International* 122: 725–734.
King SD (2009) On topography and geoid from 2D stagnant-lid convection calculations. *Geochemistry Geophysics Geosystems* 10: Q3002. http://dx.doi.org/10.1029/2008GC002250.
King SD (2001) Subduction zones: Observations and geodynamic models. *Physics of the Earth and Planetary Interiors* 127: 9–24.
King SD (2002) Geoid and topography over subduction zones: The effect of phase transformations. *Journal of Geophysical Research-Solid Earth* 107(B1). http://dx.doi.org/10.1029/2000JB000141.
King SD, Balachandar S, and Ita JJ (1997) Using eigenfunctions of the two-point correlation function to study convection with multiple phase transformations. *Geophysical Research Letters* 24: 703–706.
King SD, Gable CW, and Weinstein SA (1992) Models of convection-driven tectonic plates – A comparison of methods and results. *Geophysical Journal International* 109: 481–487.
King SD and Hager BH (1994) Subducted slabs and the geoid. 1. Numerical experiments with temperature-dependent viscosity. *Journal of Geophysical Research-Solid Earth* 99: 19843–19852.
King SD and Ita J (1995) Effect of slab rheology on mass-transport across a phase-transition boundary. *Journal of Geophysical Research-Solid Earth* 100: 20211–20222.
King SD, Lee C, van Keken PE, et al. (2010) A community benchmark for 2-D Cartesian compressible convection in the Earth's mantle. *Geophysical Journal International* 180: 73–87.
King SD and Masters G (1992) An inversion for radial viscosity structure using seismic tomography. *Geophysical Research Letters* 19: 1551–1554.
King SD, Raefsky A, and Hager BH (1990) Conman – Vectorizing a finite-element code for incompressible 2-dimensional convection in the Earths mantle. *Physics of the Earth and Planetary Interiors* 59: 195–207.

Kirby SH, Durham WB, and Stern LA (1991) Mantle phase-changes and deep-earthquake faulting in subducting lithosphere. *Science* 252: 216–225.

Kirby SH, Engdahl ER, and Denlinger R (1996) Intermediate-depth intraslab earthquakes and arc volcanism as physical expressions of crustal and uppermost mantle metamorphism in subducting slabs. In: *Subduction: Top to Bottom*, pp. 195–214. Washington, DC: American Geophysical Union.

Kita S, Okada T, Hasegawa A, Nakajima J, and Matsuzawa T (2010) Existence of interplane earthquakes and neutral stress boundary between the upper and lower planes of the double seismic zone beneath Tohoku and Hokkaido, northeastern Japan. *Tectonophysics* 496: 68–82.

Kneller EA, Long MD, and van Keken PE (2008) Olivine fabric transitions and shear wave anisotropy in the Ryukyu subduction system. *Earth and Planetary Science Letters* 268: 268–282.

Kneller EA and van Keken PE (2008) Effect of three-dimensional slab geometry on deformation in the mantle wedge: Implications for shear wave anisotropy. *Geochemistry, Geophysics, Geosystems* 9: Q01003.

Kneller EA, van Keken PE, Karato S, and Park J (2005) B-type olivine fabric in the mantle wedge: Insights from high-resolution non-Newtonian subduction zone models. *Earth and Planetary Science Letters* 237: 781–797.

Kneller EA, van Keken PE, Katayama I, and Karato S (2007) Stress, strain, and B-type olivine fabric in the fore-arc mantle: Sensitivity tests using high-resolution steady-state subduction zone models. *Journal of Geophysical Research-Solid Earth* 112: B04406.

Kohlstedt DL, Keppler H, and Rubie DC (1996) Solubility of water in the alpha, beta and gamma phases of (Mg, Fe)(2)SiO4. *Contributions to Mineralogy and Petrology* 123: 345–357.

Kohut EJ, Stern RJ, Kent AJR, et al. (2006) Evidence for adiabatic decompression melting in the Southern Mariana arc from high-Mg lavas and melt inclusions. *Contributions to Mineralogy and Petrology* 152: 201–221.

Korenaga J (2010) On the likelihood of plate tectonics on super-Earths: Does size matter? *Astrophysical Journal Letters* 725: L43–L46.

Kushiro I, Syono Y, and Akimoto SI (1968) Melting of a peridotite nodule at high pressures and high water pressures. *Journal of Geophysical Research* 73: 6023–6029.

La Rocca M, Creager KC, Galluzzo D, et al. (2009) Cascadia tremor located near plate interface constrained by S minus P wave times. *Science* 323: 620–623.

Lallemand S, Heuret A, and Boutelier D (2005) On the relationships between slab dip, back-arc stress, upper plate absolute motion, and crustal nature in subduction zones. *Geochemistry, Geophysics, Geosystems* 6: Q09006.

Lamb S (2006) Shear stresses on megathrusts: Implications for mountain building behind subduction zones. *Journal of Geophysical Research-Solid Earth* 111: B07401.

Lambeck K, Johnston P, and Nakada M (1990) Holocene glacial rebound and sea-level change in NW Europe. *Geophysical Journal International* 103: 451–468.

Larson KM, Kostoglodov V, Miyazaki S, and Santiago JAS (2007) The 2006 aseismic slow slip event in Guerrero, Mexico: New results from GPS. *Geophysical Research Letters* 34.

Lassak TM, Fouch MJ, Hall CE, and Kaminski E (2006) Seismic characterization of mantle flow in subduction systems: Can we resolve a hydrated mantle wedge? *Earth and Planetary Science Letters* 243: 632–649.

Lay T (1994) Seismological constraints on the velocity structure and fate of subducting lithospheric slabs: 25 years of progress. *Advances in Geophysics* 35(35): 1–180.

Le Pichon X, Mazzotti S, Henry P, and Hashimoto M (1998) Deformation of the Japanese Islands and seismic coupling: An interpretation based on GSI permanent GPS observations. *Geophysical Journal International* 134: 501–514.

Lebrun J-F, Lamarche G, and Collot J-Y (2003) Subduction initiation at a strike-slip plate boundary: The Cenozoic Pacific-Australian plate boundary, south of New Zealand. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2002JB002041.

Lee CY and King SD (2009) Effect of mantle compressibility on the thermal and flow structures of the subduction zones. *Geochemistry, Geophysics, Geosystems* 10: Q01006.

Lee C and King SD (2011) Dynamic buckling of subducting slabs reconciles geological and geophysical observations. *Earth and Planetary Science Letters* 312: 360–370.

Leeman WP (1996) *Boron and Other Fluid-Mobile Elements in Volcanic Arc Lavas: Implications for Subduction Processes*. Washington, DC: American Geophysical Union, pp. 269–276.

Lemoine FG, Smith DE, Kunz L, et al. (1997) The development of the NASA GSFC and NIMA joint geopotential model. In: Segawa J, Fujimoto H, and Okubo S (eds.) *International Association of Geodesy Symposia: Gravity, Geoid and Marine Geodesy*, pp. 461–469. Berlin: Springer-Verlag.

Lenardic A and Kaula WM (1993) A numerical treatment of geodynamic viscous flow problems involving the advection of material interfaces. *Journal of Geophysical Research: Solid Earth (1978–2012)* 98(B5): 8243–8260.

Lenardic A and Kaula WM (1994) Self-lubricated mantle convection – 2-Dimensional models. *Geophysical Research Letters* 21: 1707–1710.

Leng W and Zhong SJ (2008) Viscous heating, adiabatic heating and energetic consistency in compressible mantle convection. *Geophysical Journal International* 173: 693–702.

Leng W and Zhong SJ (2010) Constraints on viscous dissipation of plate bending from compressible mantle convection. *Earth and Planetary Science Letters* 297: 154–164.

Lev E and Hager BH (2011) Anisotropic viscosity changes subduction zone thermal structure. *Geochemistry Geophysics Geosystems* 12: Q04009. http://dx.doi.org/10.1029/2010GC003382.

Levin V, Droznin D, Park J, and Gordeev E (2004) Detailed mapping of seismic anisotropy with local shear waves in southeastern Kamchatka. *Geophysical Journal International* 158: 1009–1023.

Li C, van der Hilst RD, Engdahl ER, and Burdick S (2008) A new global model for P wave speed variations in Earth's mantle. *Geochemistry, Geophysics, Geosystems* 9: Q05018.

Lithgow-Bertelloni C and Richards MA (1995) Cenozoic plate driving forces. *Geophysical Research Letters* 22: 1317–1320.

Liu YJ and Rice JR (2005) Aseismic slip transients emerge spontaneously in three-dimensional rate and state modeling of subduction earthquake sequences. *Journal of Geophysical Research-Solid Earth* 110: B08307.

Liu YJ and Rice JR (2007) Spontaneous and triggered aseismic deformation transients in a subduction fault model. *Journal of Geophysical Research-Solid Earth* 112: B09404.

Long MD (2013) Constraints on subduction geodynamics from seismic anisotropy. *Reviews of Geophysics* 51(1): 76–112.

Long MD and Silver PG (2008) The subduction zone flow field from seismic anisotropy: A global view. *Science* 319: 315–318.

Long MD and Silver PG (2009) Shear wave splitting and mantle anisotropy: Measurements, interpretations, and new directions. *Surveys in Geophysics* 30: 407–461.

Long MD and van der Hilst RD (2005) Upper mantle anisotropy beneath Japan from shear wave splitting. *Physics of the Earth and Planetary Interiors* 151: 206–222.

Long MD and van der Hilst RD (2006) Shear wave splitting from local events beneath the Ryukyu arc: Trench-parallel anisotropy in the mantle wedge. *Physics of the earth and planetary interiors* 155(3): 300–312.

Long MD and Wirth EA (2013) Mantle flow in subduction systems: The mantle wedge flow field and implications for wedge processes. *Journal of Geophysical Research: Solid Earth* 118: 583–606. http://dx.doi.org/10.1002/jgrb.50063.

Lowman JP and Jarvis GT (1996) Continental collisions in wide aspect ratio and high Rayleigh number two-dimensional mantle convection models. *Journal of Geophysical Research-Solid Earth* 101: 25485–25497.

Magee ME and Zoback MD (1993) Evidence for a weak interplate thrust-fault along the northern Japan subduction zone and implications for the mechanics of thrust faulting and fluid expulsion. *Geology* 21: 809–812.

Manea VC, Manea M, Kostoglodov V, and Sewell G (2005) Thermo-mechanical model of the mantle wedge in Central Mexican subduction zone and a blob tracing approach for the magma transport. *Physics of the Earth and Planetary Interiors* 149: 165–186.

Marton FC, Bina CR, Stein S, and Rubie DC (1999) Effects of slab mineralogy on subduction rates. *Geophysical Research Letters* 26: 119–122.

Marton FC, Shankland TJ, Rubie DC, and Xu YS (2005) Effects of variable thermal conductivity on the mineralogy of subducting slabs and implications for mechanisms of deep earthquakes. *Physics of the Earth and Planetary Interiors* 149: 53–64.

Matsumoto T and Tomoda Y (1983) Numerical-simulation of the initiation of subduction at the fracture-zone. *Journal of Physics of the Earth* 31: 183–194.

Matsuzawa T, Umino N, Hasegawa A, and Takagi A (1986) Upper mantle velocity structure estimated from Ps-converted wave beneath the northeastern Japan arc. *Geophysical Journal of the Royal Astronomical Society* 86: 767–787.

McCausland W, Malone S, and Johnson D (2005) Temporal and spatial occurrence of deep non-volcanic tremor: From Washington to northern California. *Geophysical Research Letters* 32: L24311.

McCrory PA, Blair JL, Waldhauser F, and Oppenheimer DH (2012) Juan de Fuca slab geometry and its relation to Wadati–Benioff zone seismicity. *Journal of Geophysical Research-Solid Earth* 117: B09306.

Mckenzie DP (1969) Speculations on consequences and causes of plate motions. *Geophysical Journal of the Royal Astronomical Society* 18: 1–32.

Mckenzie D (1977) Surface deformation, gravity anomalies and convection. *Geophysical Journal of the Royal Astronomical Society* 48: 211–238.

Meade C and Jeanloz R (1991) Deep-focus earthquakes and recycling of water into the Earths mantle. *Science* 252: 68–72.

Miller MS, Gorbatov A, and Kennett BLN (2005) Heterogeneity within the subducting Pacific slab beneath the Izu–Bonin–Mariana arc: Evidence from tomography using 3D ray tracing inversion techniques. *Earth and Planetary Science Letters* 235: 331–342.

Miller MS, Kennett BLN, and Lister GS (2004) Imaging changes in morphology, geometry, and physical properties of the subducting Pacific plate along the Izu–Bonin–Mariana arc. *Earth and Planetary Science Letters* 224: 363–370.

Minear JW and Toksöz MN (1970) Thermal regime of a downgoing slab and new global tectonics. *Journal of Geophysical Research* 75: 1397–1419.

Minster JB and Jordan TH (1978) Present-day plate motions. *Journal of Geophysical Research* 83: 5331–5354.

Mitrovica JX (1996) Haskell [1935] revisited. *Journal of Geophysical Research-Solid Earth* 101: 555–569.

Mitrovica JX and Forte AM (1997) Radial profile of mantle viscosity: Results from the joint inversion of convection and postglacial rebound observables. *Journal of Geophysical Research-Solid Earth* 102: 2751–2769.

Molnar P and England P (1990) Temperatures, heat-flux, and frictional stress near major thrust faults. *Journal of Geophysical Research-Solid Earth and Planets* 95: 4833–4856.

Molnar P, Freedman D, and Shih JSF (1979) Lengths of intermediate and deep seismic zones and temperatures in downgoing slabs of lithosphere. *Geophysical Journal of the Royal Astronomical Society* 56: 41–54.

Moore JC and Saffer D (2001) Updip limit of the seismogenic zone beneath the accretionary prism of southwest Japan: An effect of diagenetic to low-grade metamorphic processes and increasing effective stress. *Geology* 29: 183–186.

Moresi L and Gurnis M (1996) Constraints on the lateral strength of slabs from three-dimensional dynamic flow models. *Earth and Planetary Science Letters* 138: 15–28.

Moresi LN and Solomatov VS (1995) Numerical investigation of 2D convection with extremely large viscosity variations. *Physics of Fluids* 7: 2154–2162.

Morgan WJ (1968) Rises trenches great faults and crustal blocks. *Journal of Geophysical Research* 73: 1959–1982.

Morris PA (1995) Slab melting as an explanation of quaternary volcanism and aseismicity in southwest Japan. *Geology* 23: 395–398.

Morris JD, Leeman WP, and Tera F (1990) The subducted component in island-arc lavas – Constraints from Be isotopes and B–Be systematics. *Nature* 344: 31–36.

Moucha R, Forte AM, Rowley DB, Mitrovica JX, Simmons NA, and Grand SP (2008) Mantle convection and the recent evolution of the Colorado Plateau and the Rio Grande rift valley. *Geology* 36(6): 439–442.

Moucha R, Forte AM, Rowley DB, Mitrovica JX, Simmons NA, and Grand SP (2009) Deep mantle forces and the uplift of the Colorado Plateau. *Geophysical Research Letters* 36. http://dx.doi.org/10.1029/2009GL016002.

Moyen JF and van Hunen J (2012) Short-term episodicity of Archaean plate tectonics. *Geology* 40: 451–454.

Muller RD, Roest WR, Royer JY, Gahagan LM, and Sclater JG (1997) Digital isochrons of the world's ocean floor. *Journal of Geophysical Research-Solid Earth* 102: 3211–3214.

Muller RD, Sdrolias M, Gaina C, and Roest WR (2008) Age, spreading rates, and spreading asymmetry of the world's ocean crust. *Geochemistry, Geophysics, Geosystems* 9: Q04006.

Nakagawa T and Tackley PJ (2004) Thermo-chemical structure in the mantle arising from a three-component convective system and implications for geochemistry. *Physics of the Earth and Planetary Interiors* 146: 125–138.

Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2009) Incorporating self-consistently calculated mineral physics into thermochemical mantle convection simulations in a 3-D spherical shell and its influence on seismic anomalies in Earth's mantle. *Geochemistry, Geophysics, Geosystems* 10: Q03004.

Nakajima J and Hasegawa A (2004) Shear-wave polarization anisotropy and subduction-induced flow in the mantle wedge of northeastern Japan. *Earth and Planetary Science Letters* 225: 365–377.

Nakajima J and Hasegawa A (2006) Anomalous low-velocity zone and linear alignment of seismicity along it in the subducted Pacific slab beneath Kanto, Japan: Reactivation of subducted fracture zone? *Geophysical Research Letters* 33: L16309. http://dx.doi.org/10.1029/2006GL026773.

Nakajima J, Tsuji Y, and Hasegawa A (2009) Seismic evidence for thermally-controlled dehydration reaction in subducting oceanic crust. *Geophysical Research Letters* 36: L03303.

Nakakuki T, Tagawa M, and Iwase Y (2010) Dynamical mechanisms controlling formation and avalanche of a stagnant slab. *Physics of the Earth and Planetary Interiors* 183: 309–320.

Nakiboglu SM (1982) Hydrostatic theory of the Earth and its mechanical implications. *Physics of the Earth and Planetary Interiors* 28: 302–311.

Nikolaeva K, Gerya TV, and Marques FO (2010) Subduction initiation at passive margins: Numerical modeling. *Journal of Geophysical Research-Solid Earth* 115: B03406.

Obara K (2002) Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* 296: 1679–1681.

Obara K, Hirose H, Yamamizu F, and Kasahara K (2004) Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone. *Geophysical Research Letters* 31: L23602.

Obayashi M, Sugioka H, Yoshimitsu J, and Fukao Y (2006) High temperature anomalies oceanward of subducting slabs at the 410-km discontinuity. *Earth and Planetary Science Letters* 243: 149–158.

Obayashi M, Yoshimitsu J, and Fukao Y (2009) Tearing of stagnant slab. *Science* 324: 1173–1175.

O'Driscoll LJ, Richards MA, and Humphreys ED (2012) Nazca–South America interactions and the late Eocene–late Oligocene flat-slab episode in the central Andes. *Tectonics* 31:(TC2013). http://dx.doi.org/10.1029/2011TC003036.

Ogawa M (1987) Shear instability in a viscoelastic material as the cause of deep-focus earthquakes. *Journal of Geophysical Research-Solid Earth and Planets* 92: 13801–13810.

Ohkura T (2000) Structure of the upper part of the Philippine Sea plate estimated by later phases of upper mantle earthquakes in and around Shikoku, Japan. *Tectonophysics* 321: 17–36.

Ohta Y, Freymueller JT, Hreinsdottir S, and Suito H (2006) A large slow slip event and the depth of the seismogenic zone in the south central Alaska subduction zone. *Earth and Planetary Science Letters* 247: 108–116.

Okal EA and Kirby SH (1998) Deep earthquakes beneath the Fiji Basin, SW Pacific: Earth's most intense deep seismicity in stagnant slabs. *Physics of the Earth and Planetary Interiors* 109: 25–63.

Oxburgh ER and Turcotte DL (1968) Problem of high heat flow and volcanism associated with zones of descending mantle convective flow. *Nature* 218: 1041–1043.

Ozawa S, Murakami M, Kaidzu M, et al. (2002) Detection and monitoring of ongoing aseismic slip in the Tokai region, Central Japan. *Science* 298: 1009–1012.

Pacheco JF, Sykes LR, and Scholz CH (1993) Nature of seismic coupling along simple plate boundaries of the subduction type. *Journal of Geophysical Research-Solid Earth* 98: 14133–14159.

Parai R and Mukhopadhyay S (2012) How large is the subducted water flux? New constraints on mantle regassing rates. *Earth and Planetary Science Letters* 317: 396–406.

Pavlis NK, Holmes SA, Kenyon SC, and Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research-Solid Earth* 117: 4406.

Pawley AR and Holloway JR (1993) Water sources for subduction zone volcanism – New experimental constraints. *Science* 260: 664–667.

Pawley AR and Wood BJ (1995) The high-pressure stability of talc and 10 Angstrom phase – Potential storage sites for H_2O in subduction zones. *American Mineralogist* 80: 998–1003.

Payero JS, Kostoglodov V, Shapiro N, et al. (2008) Nonvolcanic tremor observed in the Mexican subduction zone. *Geophysical Research Letters* 35: L07305.

Peacock SM (1991) Numerical-simulation of subduction zone pressure temperature time paths – Constraints on fluid production and arc magmatism. *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* 335: 341–353.

Peacock SM (1992) Blueschist-facies metamorphism, shear heating, and P–T–t paths in subduction shear zones. *Journal of Geophysical Research-Solid Earth* 97: 17693–17707.

Peacock SM (1993) Large-scale hydration of the lithosphere above subducting slabs. *Chemical Geology* 108: 49–59.

Peacock S (1996) Thermal and petrologic structure of subduction zones. In: Bebout GE, Scholl D, Kirby S, and Platt J (eds.) *Subduction: Top to Bottom*, pp. 119–133. Washington, DC: American Geophysical Union.

Peacock SM (2001) Are the lower planes of double seismic zones caused by serpentine dehydration in subducting oceanic mantle? *Geology* 29: 299–302.

Peacock S (2003) Thermal structure and metamorphic evolution of subducting slabs. In: Eiler JM (ed.) *Inside the Subduction Factory*, pp. 7–22. Washington, DC: American Geophysical Union.

Peacock SM, Rushmer T, and Thompson AB (1994) Partial melting of subducting oceanic-crust. *Earth and Planetary Science Letters* 121: 227–244.

Peacock SM, van Keken PE, Holloway SD, et al. (2005) Thermal structure of the Costa Rica–Nicaragua subduction zone. *Physics of the Earth and Planetary Interiors* 149: 187–200.

Peacock SM and Wang K (1999) Seismic consequences of warm versus cool subduction metamorphism: Examples from southwest and northeast Japan. *Science* 286: 937–939.

Pekeris CL (1935) Thermal convection in the interior of the Earth. *Monthly Notices of the Royal Astronomical Society Geophysics Supplement* 3: 343–367.

Peyton V, Levin V, Park J, et al. (2001) Mantle flow at a slab edge: Seismic anisotropy in the Kamchatka region. *Geophysical Research Letters* 28: 379–382.

Plank T, Cooper LB, and Manning CE (2009) Emerging geothermometers for estimating slab surface temperatures. *Nature Geoscience* 2: 611–615.

Polet J, Silver PG, Beck S, et al. (2000) Shear wave anisotropy beneath the Andes from the BANJO, SEDA, and PISCO experiments. *Journal of Geophysical Research-Solid Earth* 105: 6287–6304.

Poli S and Schmidt MW (2002) Petrology of subducted slabs. *Annual Review of Earth and Planetary Sciences* 30: 207–235.

Powell R and Holland TJB (1988) An internally consistent dataset with uncertainties and correlations. 3. Applications to geobarometry, worked examples and a computer-program. *Journal of Metamorphic Geology* 6: 173–204.

Powell R, Holland T, and Worley B (1998) Calculating phase diagrams involving solid solutions via non-linear equations, with examples using THERMOCALC. *Journal of Metamorphic Geology* 16: 577–588.

Pozgay SH, Wiens DA, Conder JA, Shiobara H, and Sugioka H (2007) Complex mantle flow in the Mariana subduction system: Evidence from shear wave splitting. *Geophysical Journal International* 170: 371–386.

Puster P, Hager BH, and Jordan TH (1995) Mantle convection experiments with evolving plates. *Geophysical Research Letters* 22: 2223–2226.

Quinteros J, Sobolev SV, and Popov AA (2010) Viscosity in transition zone and lower mantle: Implications for slab penetration. *Geophysical Research Letters* 37: L09307.

Ranero CR, Morgan JP, McIntosh K, and Reichert C (2003) Bending-related faulting and mantle serpentinization at the Middle America trench. *Nature* 425: 367–373.

Ratcliff JT, Schubert G, and Zebib A (1995) 3-Dimensional variable viscosity convection of an infinite Prandtl number Boussinesq fluid in a spherical-shell. *Geophysical Research Letters* 22: 2227–2230.

Ratcliff JT, Schubert G, and Zebib A (1996) Effects of temperature-dependent viscosity on thermal convection in a spherical shell. *Physica D* 97: 242–252.

Regenauer-Lieb K and Yuen DA (2000) Fast mechanisms for the formation of new plate boundaries. *Tectonophysics* 322: 53–67.

Regenauer-Lieb K and Yuen DA (2003) Modeling shear zones in geological and planetary sciences: Solid- and fluid-thermal–mechanical approaches. *Earth-Science Reviews* 63: 295–349.

Regenauer-Lieb K, Yuen DA, and Branlund J (2001) The initiation of subduction: Criticality by addition of water? *Science* 294: 578–580.

Ribe NM (2003) Periodic folding of viscous sheets. *Physical Review E* 68: 036305.

Ribe NM, Stutzmann E, Ren Y, and van der Hilst R (2007) Buckling instabilities of subducted lithosphere beneath the transition zone. *Earth and Planetary Science Letters* 254: 173–179.

Ricard Y, Fleitout L, and Froidevaux C (1984) Geoid heights and lithospheric stresses for a dynamic Earth. *Annales Geophysicae* 2: 267–286.

Ricard Y, Vigny C, and Froidevaux C (1989) Mantle heterogeneities, geoid, and plate motion – A Monte-Carlo inversion. *Journal of Geophysical Research-Solid Earth and Planets* 94: 13739–13754.

Ricard Y and Wuming B (1991) Inferring the viscosity and the 3-D density structure of the mantle from geoid, topography and plate velocities. *Geophysical Journal International* 105: 561–571.

Richard G and Bercovici D (2009) Water induced convection in the Earth's mantle transition zone. *Journal of Geophysical Research* 114: B01205. http://dx.doi.org/10.1029/2008JB005734.

Richard G, Bercovici D, and Karato S (2006) Hydration of the mantle transition zone by subducting slabs. *Earth and Planetary Science Letters* 251: 156–167.

Richards MA and Engebretson DC (1992) Large-scale mantle convection and the history of subduction. *Nature* 355: 437–440.

Richards MA and Hager BH (1984) Geoid anomalies in a dynamic Earth. *Journal of Geophysical Research* 89: 5987–6002.

Richards MA and Hager BH (1989) Effects of lateral viscosity variations on long-wavelength geoid anomalies and topography. *Journal of Geophysical Research-Solid Earth and Planets* 94: 10299–10313.

Richter FM (1973) Dynamical models for sea-floor spreading. *Transactions-American Geophysical Union* 54: 216.

Richter FM (1977) Driving mechanism of plate tectonics. *Tectonophysics* 38: 61–88.

Riedel MR and Karato S (1997) Grain-size evolution in subducted oceanic lithosphere associated with the olivine-spinel transformation and its effects on rheology. *Earth and Planetary Science Letters* 148: 27–43.

Ringwood AE (1975) *Composition and Petrology of the Earth's Mantle*. New York: Mc Graw-Hill.

Ringwood AE and Irifune T (1988) Nature of the 650-km seismic discontinuity – Implications for mantle dynamics and differentiation. *Nature* 331: 131–136.

Ritsema J, van Heijst HJ, and Woodhouse JH (2004) Global transition zone tomography. *Journal of Geophysical Research-Solid Earth* 109: B02302.

Ritzert M and Jacoby WR (1992) Geoid effects in a convecting system with lateral viscosity variations. *Geophysical Research Letters* 19: 1547–1550.

Rogers G and Dragert H (2003) Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science* 300: 1942–1943.

Rondenay S, Abers GA, and Van Keken PE (2008) Seismic imaging of subduction zone metamorphism. *Geology* 36: 275–278.

Ruff LJ and Tichelaar BW (1996) What controls the seismogenic plate interface in subduction zones? In: *Subduction: Top to Bottom*, pp. 105–111. Washington, DC: American Geophysical Union.

Runcorn SK (1967) Flow in mantle inferred from low degree harmonics of geopotential. *Geophysical Journal of the Royal Astronomical Society* 14: 375–384.

Rupke LH, Morgan JP, Hort M, and Connolly JAD (2004) Serpentine and the subduction zone water cycle. *Earth and Planetary Science Letters* 223: 17–34.

Russo RM and Silver PG (1994) Trench-parallel flow beneath the Nazca plate from seismic anisotropy. *Science* 263: 1105–1111.

Saffer DM and Marone C (2003) Comparison of smectite- and illite-rich gouge frictional properties: Application to the updip limit of the seismogenic zone along subduction megathrusts. *Earth and Planetary Science Letters* 215: 219–235.

Saffer DM and Tobin HJ (2011) Hydrogeology and mechanics of subduction zone forearcs: Fluid flow and pore pressure. *Annual Review of Earth and Planetary Sciences* 39(39): 157–186.

Savage MK (1999) Seismic anisotropy and mantle deformation: What have we learned from shear wave splitting? *Reviews of Geophysics* 37: 65–106.

Schellart WP (2004a) Kinematics of subduction and subduction-induced flow in the upper mantle. *Journal of Geophysical Research-Solid Earth* 109: B07401.

Schellart WP (2004b) Quantifying the net slab pull force as a driving mechanism for plate tectonics. *Geophysical Research Letters* 31: L07611.

Schellart WP, Freeman J, Stegman DR, Moresi L, and May D (2007) Evolution and diversity of subduction zones controlled by slab width. *Nature* 446: 308–311.

Schmeling H, Babeyko AY, Enns A, et al. (2008) A benchmark comparison of spontaneous subduction models—Towards a free surface. *Physics of the Earth and Planetary Interiors* 171: 198–223.

Schmeling H, Monz R, and Rubie DC (1999) The influence of olivine metastability on the dynamics of subduction. *Earth and Planetary Science Letters* 165: 55–66.

Schmidt MW and Poli S (1998) Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth and Planetary Science Letters* 163: 361–379.

Schmidt M and Poli S (2003) Petrology of subducted slabs. *Earth and Planetary Sciences* 30: 207–235.

Scholz CH (1990) *The Mechanics of Earthquakes and Faulting*. Cambridge: Cambridge University Press.

Scholz CH and Campos J (1995) On the mechanism of seismic decoupling and back-arc spreading at subduction zones. *Journal of Geophysical Research-Solid Earth* 100: 22103–22115.

Schubert G and Sandwell DT (1995) A global survey of possible subduction sites on Venus. *Icarus* 117: 173–196.

Schubert G and Turcotte DL (1971) Phase changes and mantle convection. *Journal of Geophysical Research* 76: 1424–1432.

Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. New York: Cambridge University Press.

Schubert G, Yuen DA, and Turcotte DL (1975) Role of phase-transitions in a dynamic mantle. *Geophysical Journal of the Royal Astronomical Society* 42: 705–735.

Schubert G and Zhang K (1997) Foundering of the lithosphere at the onset of subduction. *Geophysical Research Letters* 24: 1527–1529.

Seno T (2009) Determination of the pore fluid pressure ratio at seismogenic megathrusts in subduction zones: Implications for strength of asperities and Andean-type mountain building. *Journal of Geophysical Research-Solid Earth* 114: B05405.

Shelly DR, Beroza GC, and Ide S (2007) Non-volcanic tremor and low-frequency earthquake swarms. *Nature* 446: 305–307.

Shelly DR, Beroza GC, Ide S, and Nakamula S (2006) Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* 442: 188–191.

Silver PG (1996) Seismic anisotropy beneath the continents: Probing the depths of geology. *Annual Review of Earth and Planetary Sciences* 24: 285–432.

Silver PG, Carlson RW, and Olson P (1998) Deep slabs, geochemical heterogeneity and the large scale structure of mantle convection: Investigation of an enduring paradox. *Annual Review of Earth and Planetary Sciences* 16: 477–541.

Simmons NA, Myers SC, Johannesson G, and Matzel E (2012) LLNL-G3Dv3: Global P wave tomography model for improved regional and teleseismic travel time prediction. *Journal of Geophysical Research-Solid Earth* 117: B10302. http://dx.doi.org/10.1029/2012JB009525.

Simoes M, Avouac JP, Cattin R, and Henry P (2004) The Sumatra subduction zone: A case for a locked fault zone extending into the mantle. *Journal of Geophysical Research-Solid Earth* 109: B10402.
Skarbek RM and Saffer DM (2009) Pore pressure development beneath the decollement at the Nankai subduction zone: Implications for plate boundary fault strength and sediment dewatering. *Journal of Geophysical Research-Solid Earth* 114: B07401.
Skora S and Blundy J (2010) High-pressure hydrous phase relations of radiolarian clay and implications for the involvement of subducted sediment in arc magmatism. *Journal of Petrology* 51: 2211–2243.
Smith PM and Asimow PD (2005) Adiabat_1ph: A new public front-end to the MELTS, pMELTS, and pHMELTS models. *Geochemistry, Geophysics, Geosystems* 6: Q02004.
Smith GP, Wiens DA, Fischer KM, et al. (2001) A complex pattern of mantle flow in the Lau backarc. *Science* 292: 713–716.
Smyth JR (1994) A crystallographic model for hydrous wadsleyite (Beta-Mg_2SiO_4) – An ocean in the Earths interior. *American Mineralogist* 79: 1021–1024.
Smyth JR, Kawamoto T, Jacobsen SD, et al. (1997) Crystal structure of monoclinic hydrous wadsleyite [beta-(Mg, Fe)(2)SiO4]. *American Mineralogist* 82: 270–275.
Solomatov VS (1995) Scaling of temperature-dependent and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.
Solomatov VS and Moresi LN (1996) Stagnant lid convection on Venus. *Journal of Geophysical Research-Planets* 101: 4737–4753.
Song T-RA and Kim Y (2012) Anisotropic uppermost mantle in young subducted slab underplating Central Mexico. *Nature Geoscience* 5: 55–59.
Spiegelman M (2003) Linear analysis of melt band formation by simple shear. *Geochemistry, Geophysics, Geosystems* 4: 8615.
Spiegelman M and Mckenzie D (1987) Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth and Planetary Science Letters* 83: 137–152.
Spinelli GA and Wang K (2008) Effects of fluid circulation in subducting crust on Nankai margin seismogenic zone temperatures. *Geology* 36: 887–890.
Stadler G, Gurnis M, Burstedde C, et al. (2010) The dynamics of plate tectonics and mantle flow: From local to global scales. *Science* 329: 1033–1038.
Staudigel H and King SD (1992) Ultrafast subduction – The key to slab recycling efficiency and mantle differentiation. *Earth and Planetary Science Letters* 109: 517–530.
Staudigel H, Plank T, White B, and Schimincke HU (1996) Geochemical fluxes during seafloor alteration of the basaltic upper oceanic crust: DSDP sites 417 and 418. In: *Subduction: Top to Bottom*, pp. 19–38. Washington, DC: American Geophysical Union.
Stegman DR, Farrington R, Capitanio FA, and Schellart WP (2010) A regime diagram for subduction styles from 3-D numerical models of free subduction. *Tectonophysics* 483: 29–45.
Stegman DR, Freeman J, Schellart WP, Moresi L, and May D (2006) Influence of trench width on subduction hinge retreat rates in 3-D models of slab rollback. *Geochemistry, Geophysics, Geosystems* 7: 3012.
Stein C and Hansen U (2008) Plate motions and the viscosity structure of the mantle – Insights from numerical modelling. *Earth and Planetary Science Letters* 272: 29–40.
Stein C, Lowman JP, and Hansen U (2013) The influence of mantle internal heating on lithospheric mobility: Implications for super-Earth. *Earth and Planetary Science Letters* 361: 448–459.
Stein SA and Rubie DC (1999) Perspectives: Geophysics – Deep earthquakes in real slabs. *Science* 286: 909–910.
Stein C, Schmalzl J, and Hansen U (2004) The effect of rheological parameters on plate behaviour in a self-consistent model of mantle convection. *Physics of the Earth and Planetary Interiors* 142: 225–255.
Stein S and Stein CA (1996) Thermo-mechanical evolution of oceanic lithosphere: Implications for the subduction process and deep earthquakes. In: *Subduction: Top to Bottom*, pp. 1–17. Washington, DC: American Geophysical Union.
Stern RJ (2002) Subduction zones. *Reviews of Geophysics* 40: 3–38.
Stevenson DJ and Turner JS (1977) Angle of subduction. *Nature* 270: 334–336.
Syracuse EM and Abers GA (2006) Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochemistry, Geophysics, Geosystems* 7: Q05017.
Syracuse EM, Abers GA, Fischer K, et al. (2008) Seismic tomography and earthquake locations in the Nicaraguan and Costa Rican upper mantle. *Geochemistry, Geophysics, Geosystems* 9: Q07S08.
Syracuse EM, van Keken PE, and Abers GA (2010) The global range of subduction zone thermal models. *Physics of the Earth and Planetary Interiors* 183: 73–90.
Tackley PJ (1998) Self-consistent generation of tectonic plates in three-dimensional mantle convection. *Earth and Planetary Science Letters* 157: 9–22.
Tackley PJ (2000a) Mantle convection and plate tectonics: Toward an integrated physical and chemical theory. *Science* 288: 2002–2007.
Tackley PJ (2000b) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations 2. Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems* 1: 1026.
Tackley P, Ammann M, Brodholt J, Dobson D, and Valencia D (2013) Mantle dynamics in super-earths: Post-perovskite rheology and self-regulation of viscosity. *Icarus* 225(1): 50–61.
Tackley PJ, Stevenson DJ, Glatzmaier GA, and Schubert G (1993) Effects of an endothermic phase-transition at 670 km depth in a spherical model of convection in the Earths mantle. *Nature* 361: 699–704.
Tackley PJ, Stevenson DJ, Glatzmaier GA, and Schubert G (1994) Effects of multiple phase-transitions in a 3-dimensional spherical model of convection in Earths mantle. *Journal of Geophysical Research-Solid Earth* 99: 15877–15901.
Tajima F and Grand SP (1998) Variation of transition zone high-velocity anomalies and depression of 660 km discontinuity associated with subduction zones from the southern Kuriles to Izu–Bonin and Ryukyu. *Journal of Geophysical Research* 103(B7): 15015–15036. http://dx.doi.org/10.1029/98JB00752.
Tan E and Gurnis M (2005) Metastable superplumes and mantle compressibility. *Geophysical Research Letters* 32: L20307.
Tan E and Gurnis M (2007) Compressible thermochemical convection and application to lower mantle structures. *Journal of Geophysical Research-Solid Earth* 112: B06304.
Tichelaar BW and Ruff LJ (1993) Depth of seismic coupling along subduction zones. *Journal of Geophysical Research-Solid Earth* 98: 2017–2037.
Tobin HJ and Saffer DM (2009) Elevated fluid pressure and extreme mechanical weakness of a plate boundary thrust, Nankai Trough subduction zone. *Geology* 37: 679–682.
Tono Y, Fukao Y, Kunugi T, and Tsuboi S (2009) Seismic anisotropy of the Pacific slab and mantle wedge beneath the Japanese islands. *Journal of Geophysical Research-Solid Earth* 114: B07307.
Torii Y and Yoshioka S (2007) Physical conditions producing slab stagnation: Constraints of the Clapeyron slope, mantle viscosity, trench retreat, and dip angles. *Tectonophysics* 445: 200–209.
Tosi N, Čadek O, and Martinec Z (2009) Subducted slabs and lateral viscosity variations: Effects on the long-wavelength geoid. *Geophysical Journal International* 179: 813–826. http://dx.doi.org/10.1111/j.1365-246X.2009.04335.x.
Tosi N, Yuen DA, and Cadek O (2010) Dynamical consequences in the lower mantle with the post-perovskite phase change and strongly depth-dependent thermodynamic and transport properties. *Earth and Planetary Science Letters* 298: 229–243.
Toth J and Gurnis M (1998) Dynamics of subduction initiation at preexisting fault zones. *Journal of Geophysical Research-Solid Earth* 103: 18053–18067.
Trompert R and Hansen U (1998) Mantle convection simulations with rheologies that generate plate-like behaviour. *Nature* 395: 686–689.
Tsuji Y, Nakajima J, and Hasegawa A (2008) Tomographic evidence for hydrated oceanic crust of the Pacific slab beneath northeastern Japan: Implications for water transportation in subduction zones. *Geophysical Research Letters* 35: L14308.
Turcotte DL, Hsui AT, Torrance KE, and Schubert G (1974) Influence of viscous dissipation on Benard convection. *Journal of Fluid Mechanics* 64: 369–374.
Turcotte DL and Oxburgh ER (1967) Finite amplitude convective cells and continental drift. *Journal of Fluid Mechanics* 28: 29–42.
Turcotte DL and Schubert G (1982) *Geodynamics: Applications of Continuum Physics of Geological Problems*. New York, NY: John Wiley and Sons.
Ulmer P (2001) Partial melting in the mantle wedge – The role of H(2)O in the genesis of mantle-derived 'arc-related' magmas. *Physics of the Earth and Planetary Interiors* 127: 215–232.
Ulmer P and Trommsdorff V (1995) Serpentine stability to mantle depths and subduction-related magmatism. *Science* 268: 858–861.
Valencia D and O'connell RJ (2009) Convection scaling and subduction on Earth and super-Earths. *Earth and Planetary Science Letters* 286: 492–502.
Valencia D, O'Connell RJ, and Sasselov DD (2007) Inevitability of plate tectonics on super-Earths. *Astrophysical Journal* 670: L45–L48.
van den Berg AP, Yuen DA, and van Keken PE (1991) Effects of depth-variations in creep laws on the formation of plates in mantle dynamics. *Geophysical Research Letters* 18: 2197–2200.
van der Hilst R (1995) Complex morphology of subducted lithosphere in the mantle beneath the Tonga trench. *Nature* 374: 154–157.
van der Hilst RD, Engdahl ER, and Spakman W (1993) Tomographic inversion of P-data and Pp-data for aspherical mantle structure below the northwest Pacific region. *Geophysical Journal International* 115: 264–302.
van der Hilst R, Engdahl R, Spakman W, and Nolet G (1991) Tomographic imaging of subducted lithosphere below northwest Pacific island arcs. *Nature* 353: 37–43.

van der Hilst R and Seno T (1993) Effects of relative plate motion on the deep-structure and penetration depth of slabs below the Izu–Bonin and Mariana island arcs. *Earth and Planetary Science Letters* 120: 395–407.
van der Hilst RD, Widiyantoro S, and Engdahl ER (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386: 578–584.
van der Lee S, Regenauer-Lieb K, and Yuen DA (2008) The role of water in connecting past and future episodes of subduction. *Earth and Planetary Science Letters* 273: 15–27.
van Heck H and Tackley P (2011) Plate tectonics on super-Earths: Equally or more likely than on Earth. *Earth and Planetary Science Letters* 310(3–4): 252–261.
van Hunen J and Moyen JF (2012) Archean subduction: Fact or fiction? *Annual Review of Earth and Planetary Sciences* 40: 195–219.
van Hunen J, van den Berg AP, and Vlaar NJ (2000) A thermo-mechanical model of horizontal subduction below an overriding plate. *Earth and Planetary Science Letters* 182: 157–169.
van Hunen J, van den Berg AP, and Vlaar NJ (2001) Latent heat effects of the major mantle phase transitions on low-angle subduction. *Earth and Planetary Science Letters* 190: 125–135.
van Hunen J, van den Berg AP, and Vlaar NJ (2002) On the role of subducting oceanic plateaus in the development of shallow flat subduction. *Tectonophysics* 352: 317–333.
van Hunen J, van den Berg AP, and Vlaar NJ (2004) Various mechanisms to induce present-day shallow flat subduction and implications for the younger Earth: A numerical parameter study. *Physics of the Earth and Planetary Interiors* 146: 179–194.
van Keken PE (2003) The structure and dynamics of the mantle wedge. *Earth and Planetary Science Letters* 215: 323–338.
van Keken PE, Currie C, King SD, et al. (2008) A community benchmark for subduction zone modeling. *Physics of the Earth and Planetary Interiors* 171: 187–197.
van Keken PE, Hacker BR, Syracuse EM, and Abers GA (2011) Subduction factory: 4. Depth-dependent flux of H_2O from subducting slabs worldwide. *Journal of Geophysical Research-Solid Earth* 116: B01401.
van Keken PE, Kiefer B, and Peacock SM (2002) High-resolution models of subduction zones: Implications for mineral dehydration reactions and the transport of water into the deep mantle. *Geochemistry, Geophysics, Geosystems* 3: 1056.
van Keken PE and King SD (2005) Thermal structure and dynamics of subduction zones: Insights from observations and modeling. *Physics of the Earth and Planetary Interiors* 149: 1–6.
van Keken PE, Kita S, and Nakajima J (2012) Thermal structure and intermediate-depth seismicity in the Tohoku-Hokkaido subduction zones. *Solid Earth* 3: 355–364.
Vassiliou MS and Hager BH (1988) Subduction zone earthquakes and stress in slabs. *Pure and Applied Geophysics* 128: 547–624.
Vassiliou MS, Hager BH, and Raefsky A (1984) The distribution of earthquakes with depth and stress in subducting slabs. *Journal of Geodynamics* 1: 11–28.
Venig Meines FA (1954) Indonesian archipelago: A geophysical study. *Geological Society of America Bulletin* 65: 143–164.
Vlaar NJ (1983) Thermal anomalies and magmatism due to lithospheric doubling and shifting. *Earth and Planetary Science Letters* 65: 322–330.
Von Herzen R, Ruppel C, Molnar P, et al. (2001) A constraint on the shear stress at the Pacific-Australian plate boundary from heat flow and seismicity at the Kermadec forearc. *Journal of Geophysical Research-Solid Earth* 106: 6817–6833.
Vrolijk P (1990) On the mechanical role of smectite in subduction zones. *Geology* 18: 703–707.
Wada I, Behn MD, and He JH (2011) Grain-size distribution in the mantle wedge of subduction zones. *Journal of Geophysical Research-Solid Earth* 116: B10203.
Wada I, Behn MD, and Shaw AM (2012) Effects of heterogeneous hydration in the incoming plate, slab rehydration, and mantle wedge hydration on slab-derived H_2O flux in subduction zones. *Earth and Planetary Science Letters* 353: 60–71.
Wada I and Wang K (2009) Common depth of slab–mantle decoupling: Reconciling diversity and uniformity of subduction zones. *Geochemistry, Geophysics, Geosystems* 10: Q10009.
Wada I, Wang K, He JG, and Hyndman RD (2008) Weakening of the subduction interface and its effects on surface heat flow, slab dehydration, and mantle wedge serpentinization. *Journal of Geophysical Research-Solid Earth* 113: B04402.
Wallace LM, Beavan J, McCaffrey R, and Darby D (2004) Subduction zone coupling and tectonic block rotations in the North Island, New Zealand. *Journal of Geophysical Research-Solid Earth* 109: B12406.
Wallace LM, Reyners M, Cochran U, et al. (2009) Characterizing the seismogenic zone of a major plate boundary subduction thrust: Hikurangi Margin, New Zealand. *Geochemistry, Geophysics, Geosystems* 10: Q10006.
Wallmann K (2001) The geological water cycle and the evolution of marine delta O-18 values. *Geochimica et Cosmochimica Acta* 65: 2469–2485.
Wang K (2002) Unbending combined with dehydration embrittlement as a cause for double and triple seismic zones. *Geophysical Research Letters* 29: 1889.
Wang KL, Dragert H, Kao H, and Roeloffs E (2008) Characterizing an "uncharacteristic" ETS event in northern Cascadia. *Geophysical Research Letters* 35: L15303.
Wang KL and He JH (1999) Mechanics of low-stress forearcs: Nankai and Cascadia. *Journal of Geophysical Research-Solid Earth* 104: 15191–15205.
Wang KL, Mulder T, Rogers GC, and Hyndman RD (1995) Case for very-low coupling stress on the Cascadia subduction fault. *Journal of Geophysical Research-Solid Earth* 100: 12907–12918.
Wang KL and Suyehiro K (1999) How does plate coupling affect crustal stresses in Northeast and Southwest Japan? *Geophysical Research Letters* 26: 2307–2310.
Wang K, Wells R, Mazzotti S, Hyndman RD, and Sagiya T (2003) A revised dislocation model of interseismic deformation of the Cascadia subduction zone. *Journal of Geophysical Research-Solid Earth* 108: 2026.
Wang J and Zhao D (2012) P wave anisotropic tomography of the Nankai subduction zone in Southwest Japan. *Geochemistry, Geophysics, Geosystems* 13: Q05017. http://dx.doi.org/10.1029/2012GC004081.
Wech AG and Creager KC (2007) Cascadia tremor polarization evidence for plate interface slip. *Geophysical Research Letters* 34: L22306.
Weinstein SA and Olson PL (1992) Thermal convection with Non-Newtonian Plates. *Geophysical Journal International* 111: 515–530.
Wessel P and Smith WHF (1991) Free software helps map and display data. *Eos, Transactions of the American Geophysical Union* 72: 441–446.
Widiyantoro S, Kennett BLN, and van der Hilst RD (1999) Seismic tomography with P and S data reveals lateral variations in the rigidity of deep slabs. *Earth and Planetary Science Letters* 173: 91–100.
Wiens DA (2001) Seismological constraints on the mechanism of deep earthquakes: Temperature dependence of deep earthquake source properties. *Physics of the Earth and Planetary Interiors* 127: 145–163.
Wiens DA, Conder JA, and Faul UH (2008) The seismic structure and dynamics of the mantle wedge. *Annual Review of Earth and Planetary Sciences* 36: 421–455.
Wirth EA and Long MD (2010) Frequency dependent shear wave splitting beneath the Japan and Izu–Bonin subduction zones. *Physics of the Earth and Planetary Interiors* 181: 141–154.
Wirth EA and Long MD (2012) Multiple layers of seismic anisotropy and a low-velocity layer in the mantle wedge beneath Japan: Evidence from teleseismic receiver functions. *Geochemistry, Geophysics, Geosystems* 13: Q08005. http://dx.doi.org/10.1029/2012GC004180.
Wu BJ, Conrad CP, Heuret A, Lithgow-Bertelloni C, and Lallemand S (2008) Reconciling strong slab pull and weak plate bending: The plate motion constraint on the strength of mantle slabs. *Earth and Planetary Science Letters* 272: 412–421.
Yamasaki T and Seno T (2003) Double seismic zone and dehydration embrittlement of the subducting slab. *Journal of Geophysical Research-Solid Earth* 108: 2212.
Yang T, Grand SP, Wilson D, et al. (2009) Seismic structure beneath the Rivera Subduction Zone from finite frequency seismic tomography. *Journal of Geophysical Research-Solid Earth* 114. http://dx.doi.org/10.1029/2008JB005830.
Yogodzinski GM, Kay RW, Volynets ON, Koloskov AV, and Kay SM (1995) Magnesian andesite in the western Aleutian Komandorsky region: Implications. *Geological Society of America Bulletin* 107: 505–519.
Yoshioka S and Sanshadokoro H (2002) Numerical simulations of deformation and dynamics of horizontally lying slabs. *Geophysical Journal International* 151: 69–82.
Yuan X, Sobolev SV, Kind R, et al. (2000) Subduction and collision processes in the Central Andes constrained by converted seismic phases. *Nature* 408: 958–961.
Yuen DA, Fleitout L, Schubert G, and Froidevaux C (1978) Shear deformation zones along major transform faults and subducting slabs. *Geophysical Journal of the Royal Astronomical Society* 54: 93–119.
Zhang SX and Yuen DA (1996) Various influences on plumes and dynamics in time-dependent, compressible mantle convection in 3-D spherical shell. *Physics of the Earth and Planetary Interiors* 94: 241–267.
Zhao DP, Lei JS, and Tang RY (2004) Origin of the Changbai intraplate volcanism in Northeast China: Evidence from seismic tomography. *Chinese Science Bulletin* 49: 1401–1408.
Zhao DP, Matsuzawa T, and Hasegawa A (1997) Morphology of the subducting slab boundary in the northeastern Japan arc. *Physics of the Earth and Planetary Interiors* 102: 89–104.
Zhao DP, Tian Y, Lei JS, Liu LC, and Zheng SH (2009) Seismic image and origin of the Changbai intraplate volcano in East Asia: Role of big mantle wedge above the stagnant Pacific slab. *Physics of the Earth and Planetary Interiors* 173: 197–206.
Zhong SJ and Davies GF (1999) Effects of plate and slab viscosities on the geoid. *Earth and Planetary Science Letters* 170: 487–496.

Zhong SJ and Gurnis M (1992) Viscous-flow model of a subduction zone with a faulted lithosphere – Long and short wavelength topography, gravity and geoid. *Geophysical Research Letters* 19: 1891–1894.

Zhong SJ and Gurnis M (1994) Controls on trench topography from dynamic-models of subducted slabs. *Journal of Geophysical Research-Solid Earth* 99: 15683–15695.

Zhong SJ and Gurnis M (1995) Mantle convection with plates and mobile, faulted plate margins. *Science* 267: 838–843.

Zhong GF and Gurnis M (1997) Dynamic interaction between tectonic plates, subducting slabs, and the mantle. *Earth Interactions* 1: 1–18.

Zhong SJ, Gurnis M, and Moresi L (1998) Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models. *Journal of Geophysical Research-Solid Earth* 103: 15255–15268.

Zhou HW and Clayton RW (1990) P and S-wave travel time inversions for subducting slab under the island arcs of the northwest Pacific. *Journal of Geophysical Research-Solid Earth and Planets* 95: 6829–6851.

Zhu GZ, Gerya TV, Honda S, Tackley PJ, and Yuen DA (2011) Influences of the buoyancy of partially molten rock on 3-D plume patterns and melt productivity above retreating slabs. *Physics of the Earth and Planetary Interiors* 185: 112–121.

Zhu GZ, Gerya TV, Yuen DA, et al. (2009) Three-dimensional dynamics of hydrous thermal–chemical plumes in oceanic subduction zones. *Geochemistry, Geophysics, Geosystems* 10: Q11006.

This page intentionally left blank

7.10 Hotspots, Large Igneous Provinces, and Melting Anomalies

MD Ballmer, University of Hawaii, Honolulu, HI, USA; Tokyo Institute of Technology, Meguro-ku, Tokyo, Japan
PE van Keken, University of Michigan, Ann Arbor, MI, USA
G Ito, University of Hawaii, Honolulu, HI, USA

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by G Ito and P E van Keken, Volume 7, pp. 371–435, © 2007, Elsevier B.V.

7.10.1 Introduction 394
7.10.2 Characteristics 394
7.10.2.1 Volcano Chains and Age Progression 394
7.10.2.1.1 Long-lived age-progressive volcanism 394
7.10.2.1.2 Short-lived age-progressive volcanism 400
7.10.2.1.3 Non-age-progressive volcanism 401
7.10.2.1.4 Continental hotspots and melting anomalies 403
7.10.2.1.5 The hotspot reference frame 403
7.10.2.2 Topographic Swells 406
7.10.2.3 Flood Basalt Volcanism 407
7.10.2.3.1 Intraplate continental LIPs 407
7.10.2.3.2 LIPs near or on continental margins 408
7.10.2.3.3 Oceanic LIPs 410
7.10.2.3.4 Connections to hotspots 411
7.10.2.3.5 Connections to climate crises and mass extinction 411
7.10.2.4 Geochemistry and Petrology of OIB 412
7.10.2.4.1 Isotope geochemistry 412
7.10.2.4.2 Major-element geochemistry 412
7.10.2.5 Mantle Seismic Anomalies 414
7.10.2.5.1 Global seismic studies 414
7.10.2.5.2 Seismic studies of major hotspots 415
7.10.2.6 Summary of Observations 424
7.10.3 Dynamic Mechanisms and Their Implications 425
7.10.3.1 Methods 425
7.10.3.2 Generating Magmas for Volcanism 425
7.10.3.2.1 Temperature 426
7.10.3.2.2 Composition 426
7.10.3.2.3 Mantle upwelling 426
7.10.3.3 Dynamics of Mantle Upwellings 426
7.10.3.3.1 Thermal boundary layer instabilities 427
7.10.3.3.2 Thermal plumes 427
7.10.3.3.3 Thermochemical plumes 428
7.10.3.3.4 Effects of variable mantle properties on plume dynamics 429
7.10.3.3.5 Small-scale sublithospheric convection 430
7.10.3.3.6 Buoyant decompression melting 431
7.10.3.3.7 Shear-driven upwelling and viscous fingering 432
7.10.3.3.8 Vertical motion of the lithosphere 434
7.10.3.4 Swells 434
7.10.3.4.1 Swell support mechanisms 434
7.10.3.4.2 Plume models and hotspot swells 435
7.10.3.5 Large Igneous Provinces 435
7.10.3.6 The Origin of OIB Geochemical Signatures 436
7.10.3.6.1 Tracing mantle heterogeneity by isotopic signatures of OIB 436
7.10.3.6.2 Constraining source materials from major-element characteristics of OIB 437
7.10.3.6.3 The origin of geographic variations in OIB geochemistry 438
7.10.3.7 Plumes in the Context of Mantle Convection and Plate Tectonics 438
7.10.3.7.1 Connection of hotspots with thermochemical structures in the deep mantle 438
7.10.3.7.2 Interaction of plumes with plate tectonic processes 439
7.10.3.7.3 Lithospheric controls on volcanism 442
7.10.4 Conclusion 442
Acknowledgments 443
References 443

7.10.1 Introduction

Volcanism on Earth and the generation of the oceanic and continental crust predominantly occur along divergent or convergent margins. At mid-oceanic ridges, decompression of the upwelling mantle induces magmatism. At subduction zones, the influx of fluids released from the metamorphic dehydration reactions in the downgoing slab leads to partial melting of the overlying hot mantle wedge. In contrast, several forms of volcanism are less readily explained by plate tectonic processes. These include, for example, the formation of intraplate hotspot islands (such as Hawaii) and excessive volcanism at plate boundaries (such as at Iceland).

The most prominent melting anomalies typically occur on top of a broad swelling of topography and are trailed by a volcano chain that parallels plate motion. In some cases, these chains project back to massive volcanic plateaus (large igneous provinces, LIPs), suggesting that hotspot activity began with magmatic outbursts as large as any in the geologic record (Duncan and Richards, 1991; Morgan, 1972; Richards et al., 1989; White and McKenzie, 1989). These observations have led to the establishment of classical 'hotspot' theory proposed by Wilson (1963, 1973), Morgan (1971, 1972), and Crough (1978), who describe hotspots as positive thermal anomalies in the mantle that are fixed relative to the motion of the tectonic plates. These plumes cause excess seafloor elevation and feed age-progressive volcanism.

Modern dating techniques, however, recover complex age patterns for many volcano chains and reveal that intraplate volcanism cannot universally be ascribed to activity at localized hotspots (Koppers et al., 2003; McNutt et al., 1997). In addition, some hotspots are not confidently linked to anomalously high mantle temperatures (Klein and Langmuir, 1987; Langmuir et al., 1992). In turn, most hotspots are fed by melting of a source that is geochemically distinct and typically more fertile than 'normal' mantle that feeds mid-ocean ridge (MOR) melting (Hart et al., 1973; Hofmann, 1997; Jackson and Dasgupta, 2008; Schilling, 1971, 1973). Thus, the term 'melting anomaly' may be more general and appropriate to describe the topic of this chapter.

The differences between the hotspot source and the 'normal' mantle in both major elements and isotopes (Jackson and Dasgupta, 2008) suggest that volcanism at melting anomalies has origins that are at least partly decoupled from plate tectonic processes. A straightforward explanation is that magmatism is sustained by convective upwellings or plumes of unusually hot, buoyant mantle, which rise from the lower mantle (Morgan, 1971, 1972; Whitehead and Luther, 1975; Wilson, 1963, 1973), possibly through a chemically heterogeneous mantle (e.g., Richter and McKenzie, 1981). The morphology of a mantle plume, simulated in laboratory and numerical experiments, is predicted to have a large 'mushroom'-shaped head and a trailing, narrower plume stem. This geometry is similar to that expected for the formation of an LIP followed by a hotspot track (e.g., Campbell and Griffiths, 1990; Richards et al., 1989). Recent regional (e.g., Wolfe et al., 2009) and global seismic studies (e.g., Boschi et al., 2007; Montelli et al., 2006; Zhao, 2007) are providing increasingly convincing evidence for mantle plumes originating in the lower mantle. Another observational link to the deep mantle is the spatial relation of hotspot and LIP locations to the edges of large low-shear-wave-velocity provinces (LLSVPs) imaged at the base of the mantle (Boschi et al., 2007; Burke and Torsvik, 2004; Torsvik et al., 2006).

Studies of hotspots have flourished over the past few decades. Recent articles and textbooks discuss some of the classic ties of hotspots to mantle plumes (e.g., Campbell, 2007; Campbell and Kerr, 2007; Condie, 2001; Davies, 1999; Foulger and Jurdy, 2007; Foulger et al., 2006; Jackson, 1998; Schubert et al., 2001; Sleep, 2006), the role of mantle plumes in deep mantle convection and chemical transport (e.g., Deschamps et al., 2011; Jellinek and Manga, 2004; Kumagai et al., 2008), and oceanic hotspots (e.g., Ballmer et al., 2011; Hekinian et al., 2004; Ito et al., 2003). It has become apparent that only a few hotspots confidently show all of the previously mentioned characteristics of the classic plume description (Clouard and Bonneville, 2001; Courtillot et al., 2003). The term 'hotspot' itself implies a localized region of anomalously high mantle temperature, but some features that were originally called hotspots may involve little or no excess heat or large distances of coeval volcanism. To understand these discrepancies, there has been significant exploration of the interaction of mantle plumes with plate tectonic processes (e.g., Kincaid et al., 2013; Mittelstaedt et al., 2012), the dynamic effect of compositional heterogeneity (Ballmer et al., 2013b; Davaille, 1999; Kumagai et al., 2008; Lin and van Keken, 2005), and mechanisms that can feed non-hot spot volcanism (Ballmer et al., 2007; Clouard and Gerbault, 2008; Conrad et al., 2011). The progress made in the last decade motivates a comprehensive review on studies of hotspots, LIPs, and melting anomalies. We here summarize the main observations, outline the mechanisms that have been proposed to cause hotspots and other melting anomalies, and pose questions that need quantitative answers.

7.10.2 Characteristics

Guided by the classical description of hotspots, we examine four main characteristics: (1) geographic age progression along volcano chains, (2) initiation by massive flood basalt volcanism, (3) a swell of anomalously shallow topography surrounding volcanoes, and (4) basaltic volcanism with geochemical distinction from most MOR basalts (MORBs). Given the marked progress in seismic methods over the past decade, we also summarize the findings of mantle seismic structure beneath hotspots and surface melt anomalies. **Table 1** provides a compilation of the previously mentioned characteristics for the 67 hotspots and melting anomalies. **Figure 1** shows a global map of their locations with abbreviations and the main LIPs that we will discuss.

7.10.2.1 Volcano Chains and Age Progression

7.10.2.1.1 Long-lived age-progressive volcanism

At least 13 hotspot chains record volcanism enduring for more than 50 My. The Hawaiian–Emperor and the Louisville chains, for example, span thousands of kilometers across the Pacific

Table 1 Summary of characteristics

Name (abbreviation)	*Hotspot E. long., N. lat.*	*Age progression?*	*Age range*	*Swell? and width (km)*	*Connection to LIP?*	*Geochemically distinct from MORB*
Pacific						
Baja (BAJ)	−113, 27	–	–	No	No	–
Bowie–Kodiak (BOW)	−130, 49.5	Ok	0.1–23.8 Ma	Yes/250 km	No	Maybe $^{206}Pb/^{204}Pb$
Caroline (CAR)	−163, 5.3	Weak	1.4 Ma (east) to 4.7–13.9 Ma (west)	–	No	No
Cobb (COB)	−128.7, 43.6	Good	1.5–29.2 Ma	Yes/370	No	No
Cook (CK)	−149.5, −23.5	No	0.2–19.4 Ma	Yes/500	No	$^{206}Pb/^{204}Pb$
Austral (AU)	−140.0, −29.37	No	0–58.1 Ma	Yes/600	No	$^{206}Pb/^{204}Pb$
Easter (EAS)	−109, −27	Good	0–25.6 Ma	yes/580	Maybe Tuamotu and Mid-Pac	$^{206}Pb/^{204}Pb$
Foundation (FOU)	−111, −39	Good	2.1–21 Ma	Yes/250	No	$^{206}Pb/^{204}Pb$
Galápagos (GAL)	−91.6, −0.4	Yes	0–14.5 Ma offshore; 69–139 Ma, Caribbean LIP	Yes/300	Caribbean LIP	$^{206}Pb/^{204}Pb$
Geologist (GEO)	−157, 19	No	82.7–84.6 Ma	–	No	–
Guadalupe (GUA)	−118, 29	–	$<$3.4 to $\sim$20.3 Ma	Maybe/?	No	–
Hawaii–Emperor (HAW)	−155.3, 18.9	Good	0–75.8 Ma	Yes/920	No	$^{3}He/^{4}He$ and $^{87}Sr/^{86}Sr$ for Hawaiian Islands but not Emperor Seamounts
Japanese–Wake (JWK)	–	No	78.6–119.7 Ma	No	No	$^{206}Pb/^{204}Pb$
Juan Fernandez (JFE)	−79, −34	Weak	1–4 Ma (2 volcanoes dated)	Yes/?	No	$^{3}He/^{4}He$ and $^{87}Sr/^{86}Sr$
Line Islands (LIN)	–	No	35.5–91.2 Ma.	Partially/?	Maybe Mid-Pac	–
Louisville (LOU)	−141.2, −53.6	Good	1.1–77.3 Ma	Yes/540	Doubtfully OJP	$^{206}Pb/^{204}Pb$, maybe $^{87}Sr/^{86}Sr$
Magellan Seamounts (MAG)	–	No	87–118.6 Ma	No	No	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Marquesas (MQS)	−138.5, −11	Ok	0.8–5.5 Ma	Yes/850	Maybe Shatsky or Hess	$^{87}Sr/^{86}Sr$, maybe $^{206}Pb/^{204}Pb$
Marshall Islands (MI)	–	No	68–138 Ma	Maybe/?	No	$^{206}Pb/^{204}Pb$
Mid-Pacific Mountains (MPM)	–	No	73.5–128 Ma	No	It could be an LIP	–
Musician (MUS)	–	Ok	65.5–95.8 Ma	No	No	–
Pitcairn (PIT)	−129.4, −25.2	Good	0–11.1 Ma	Yes/570	No	$^{87}Sr/^{86}Sr$
Pukapuka (PUK)	−165.5, −10.5	Ok	5.6–27.5 Ma	Yes/?	No	$^{206}Pb/^{204}Pb$
Samoa (SAM)	−169, −14.3	Weak	0–23 Ma	400	No	$^{87}Sr/^{86}Sr$, $^{3}He/^{4}He$, and $^{206}Pb/^{204}Pb$
San Felix (SF)	−80, −26	–	–	Yes/?	No	–
Shatsky (SHA)	–	Yes	128–145 Ma	No	It is an LIP	No

(*Continued*)

Table 1 (Continued)

Name (abbreviation)	*Hotspot E. long., N. lat.*	*Age progression?*	*Age range*	*Swell? and width (km)*	*Connection to LIP?*	*Geochemically distinct from MORB*
Society (SOC)	−148, −18	Good	0.01–4.6 Ma	Yes/?	No	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Socorro (SCR)	−111, 19	–	–	Yes/?	No	–
Tarava (TAR)	173, 3	Weak	35.9 Ma and 43.5 Ma	Yes/?	No	–
Tuamotu (TUA)	–	Good	58–74 Ma	Yes/?	No	–
North America						
Yellowstone (YEL)	−111, 44.8	Yes	16–17 Ma	Yes/600	Maybe Columbia River basalts	–
Australia						
Balleny (BAL)	164.7, −67.4	Weak	–	–	Maybe Lord Howe Rise	$^{206}Pb/^{204}Pb$ (2 analyses)
East Australia (AUS)	143, −38	–	–	–	–	–
Lord Howe (LHO)	159, −31	–	–	–	It could be LIP	–
Tasmantid (TAS)	153, −41	Yes	–	Yes/300	Maybe Lord Howe Rise	–
Africa						
Afar (AF)	42, 12	No	–	–	–	–
East Africa/ Lake Victoria (EAF)	34, 6	No	–	–	–	–
Darfur (DAR)	24,13	No	–	–	–	–
Ahaggar (HOG)	6, 23	No	–	–	–	–
Tibesti (TIB)	17, 21	No	–	–	–	–
Ascension/ Circe (ASC)	−14, −8	–	<1 Ma (Ascension) and 6 Ma (Circe)	Yes/800	No	$^{206}Pb/^{204}Pb$
Azores (AZO)	−28, 38	Seafloor spreading	0–20 Ma, possibly ~85 Ma	Yes/2300	No	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Bermuda (BER)	−65, 32	–	–	Yes/500 × 700 (parallel × perp to plate motion)	No	–
Bouvet (BOU)	3.4, −54.4	–	?	Yes/900	Agulhas Plateau?	$^{206}Pb/^{204}Pb$, maybe $^{87}Sr/^{86}Sr$ and $^{3}He/^{4}He$
Cameroon (CAM)	6, −1	No	1–32 Ma	Yes/500–600	No	$^{206}Pb/^{204}Pb$
Canaries (CAN)	−17, 28	Ok	0–68 Ma	No	No	$^{206}Pb/^{204}Pb$
Cape Verde (CAP)	−24, 15	No	Neogene	Yes/800	No	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Discovery (DIS)	−6.45, −44.45	–	35–41 Ma	Yes/600	No	–
Fernando Do Noronha (FER)	−32, −4	–	–	Yes/200–300	–	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Great Meteor (GM)	–	–	–	Yes/800	–	–
Iceland (ICE)	−17.6, 64.6	Yes	0–62 Ma	Yes/2700	N. Atlantic LIP	$^{3}He/^{4}He$
Madeira (MAD)	−17.5, 32.7	Yes	0–67 Ma	No	No	$^{206}Pb/^{204}Pb$
New England (NEW)	−57.5, 35	Yes	81–103, 122–124 Ma	No	No	$^{206}Pb/^{204}Pb$, maybe $^{87}Sr/^{86}Sr$
Shona–Agulhas (SHO)	−4, −52	–	2.5–81 Ma	~900	–	–

(*Continued*)

Table 1 (Continued)

Name (abbreviation)	*Hotspot E. long., N. lat.*	*Age progression?*	*Age range*	*Swell? and width (km)*	*Connection to LIP?*	*Geochemically distinct from MORB*
Sierra Leone (SL)	−29, 1	–	Not dated	–	It could be an LIP	–
St. Helena (SHE)	−10, −17	Yes	3–81 Ma	Yes/720	No	$^{206}Pb/^{204}Pb$
Trindade–Martim Vaz (TRN)	−12.2, −37.5	Probably	<1 Ma to ~85 Ma	Yes/1330	Small eruptions north of Paraná flood basalts and onto Brazilian margin	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Tristan–Gough (TRI)	−9.9, −40.4	Yes	0.5–80 Ma and 130 Ma	Yes/850	Rio Grande–Walvis and Paraná–Etendeka	$^{87}Sr/^{86}Sr$
Vema (VEM)	16, −32	–	>11 Ma	Yes/200–300		–
Amsterdam–St. Paul (AMS)	77, −37	No	–	yes/300–500	Maybe Kerguelen	Maybe $^{87}Sr/^{86}Sr$
Comores (COM)	44, −12	Yes	0–5.4 Ma on island chain and ~50 Ma (Seychelles)	yes/700–800	–	$^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$
Conrad (CON)	48, −54	–	Not dated	Half width 400 south of seamounts	It could be an LIP	–
Crozet (CRO)	50, −46	–	Not dated	Yes/1120	Maybe Madagascar	$^{206}Pb/^{204}Pb$, maybe $^{87}Sr/^{86}Sr$ (but few samples)
Kerguelen (KER)	63, −49	Yes	0.1–114 Ma (Kerguelen) and 38–82 Ma (Ninetyeast–Broken Ridge)	Yes/1310	It is an LIP	$^{87}Sr/^{86}Sr$
Marion (MAR)	37.8, −46.8	Weak	<0.5 Ma (Marion) and 88 Ma (Madagascar)	Half width 500 or along axis >1700	Madagascar Plateau and Madagascar island flood basalts	Maybe $^{87}Sr/^{86}Sr$
Réunion (REU)	55.5, −21	Yes	0–70 Ma	Yes/1380	Mascarene, Chagos–Laccadive, and Deccan 30–70 Ma	$^{3}He/^{4}He$ and $^{87}Sr/^{86}Sr$
Eurasia						
Eifel (EIF)	7, 50	No	–	–	–	–

Blanks indicate that there are no data, where we did not find any data, where available data are inconclusive.

Basin (~6000 and >4000 km, respectively), record volcanism starting before 75 Ma (Duncan and Clague, 1985; Duncan and Keller, 2004; Koppers et al., 2004; Watts et al., 1988), and were among the first chains that led to the establishment of the hotspot hypothesis. As both chains terminate at subduction zones, the existing volcanoes likely record only part of the activities of these hotspots (e.g., Portnyagin et al., 2009).

In addition to Hawaii and Louisville, the Galápagos is the third Pacific hotspot with a similar duration. Its interaction with the Galápagos Spreading Center has produced two chains: the Galápagos archipelago–Carnegie Ridge on the Nazca Plate (Sinton et al., 1996) and the Cocos Ridge on the Cocos Plate. The Cocos Ridge records oceanic volcanism for ~14.5 My (Werner et al., 1999) and projects toward the Caribbean LIP (Duncan and Hargraves, 1984) with $^{40}Ar/^{39}Ar$ dates of 69–139 Ma (e.g., Hoernle et al., 2004; Sinton et al., 1997). The geochemical similarity of these lavas with the Galápagos archipelago is a compelling evidence for a total life span for the Galápagos melting anomaly of ~139 My (Hoernle et al., 2002, 2004).

In the Indian Ocean, at least four hotspots have been related to long-term activity. Muller et al.'s (1993b) compilation of ages associates the Réunion hotspot to volcanism on the Mascarene Plateau at 45 Ma (Duncan et al., 1990), the Cocos–Laccadive Plateau at ~60 Ma (Duncan, 1978, 1991), and finally the Deccan flood basalts in India with dates as old as ~70 Ma (see also Sheth, 2005). The Comoros hotspot can be linked to volcanism around the Seychelles islands at 63 Ma (Emerick and Duncan, 1982; Müller et al., 1993b). Volcanism on Marion (<0.5 Ma; cf. McDougall et al., 2001) projects along

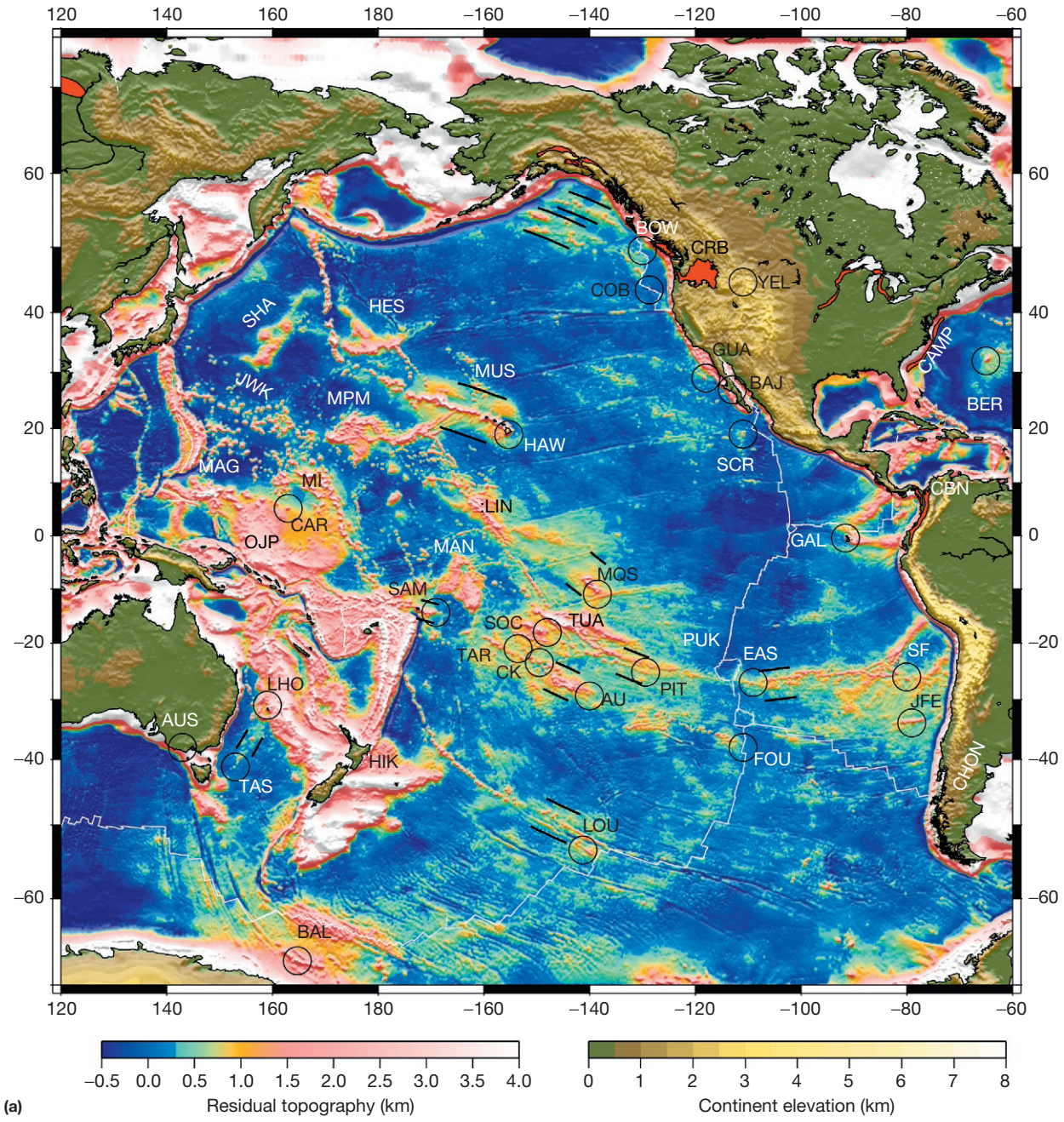


Figure 1 (Continued)

a volcanic ridge to Madagascar (Meert and Tamrat, 2006). While geologic dating is sparse, Storey et al. (1997) inferred an age progression along this track back to ~88 Ma. The Kerguelen hotspot is linked to Broken Ridge and Ninetyeast Ridge on the Australian Plate and multiple stages of volcanism on the Kerguelen Plateau dating to 114 Ma (Frey et al., 2000; Krishna et al., 2012; Nicolaysen et al., 2000; **Figure 2**).

In the Atlantic Ocean, at least seven long-lived melting anomalies have been documented. The Tristan–Gough and St. Helena chains record volcanism on the African Plate for ~80 My (O'Connor and Duncan, 1990; O'Connor and Roex, 1992; O'Connor et al., 1999). For Tristan–Gough, a connection to the Paraná flood basalts in South America and the Etendeka basalts in Namibia may even extend the duration of the melting anomaly to ~130 My (O'Connor et al., 2012; Peate, 1997). O'Connor et al. (2012) further proposed a connection of the Shona seamounts with the Agulhas Ridge to form another (zigzag-shaped) long-lived hotspot trail in the South Atlantic (~80 My). The Trindade–Martim Vaz chain on the South American Plate (Fodor and Hanan, 2000) extends to the Alto Paranaiba and Poxoreu volcanic provinces of ages ~85 Ma (Gibson et al., 1997). In the North Atlantic, the Madeira and Canary chains have recorded age-progressive volcanism for nearly 70 My (Geldmacher et al., 2005; Guillou

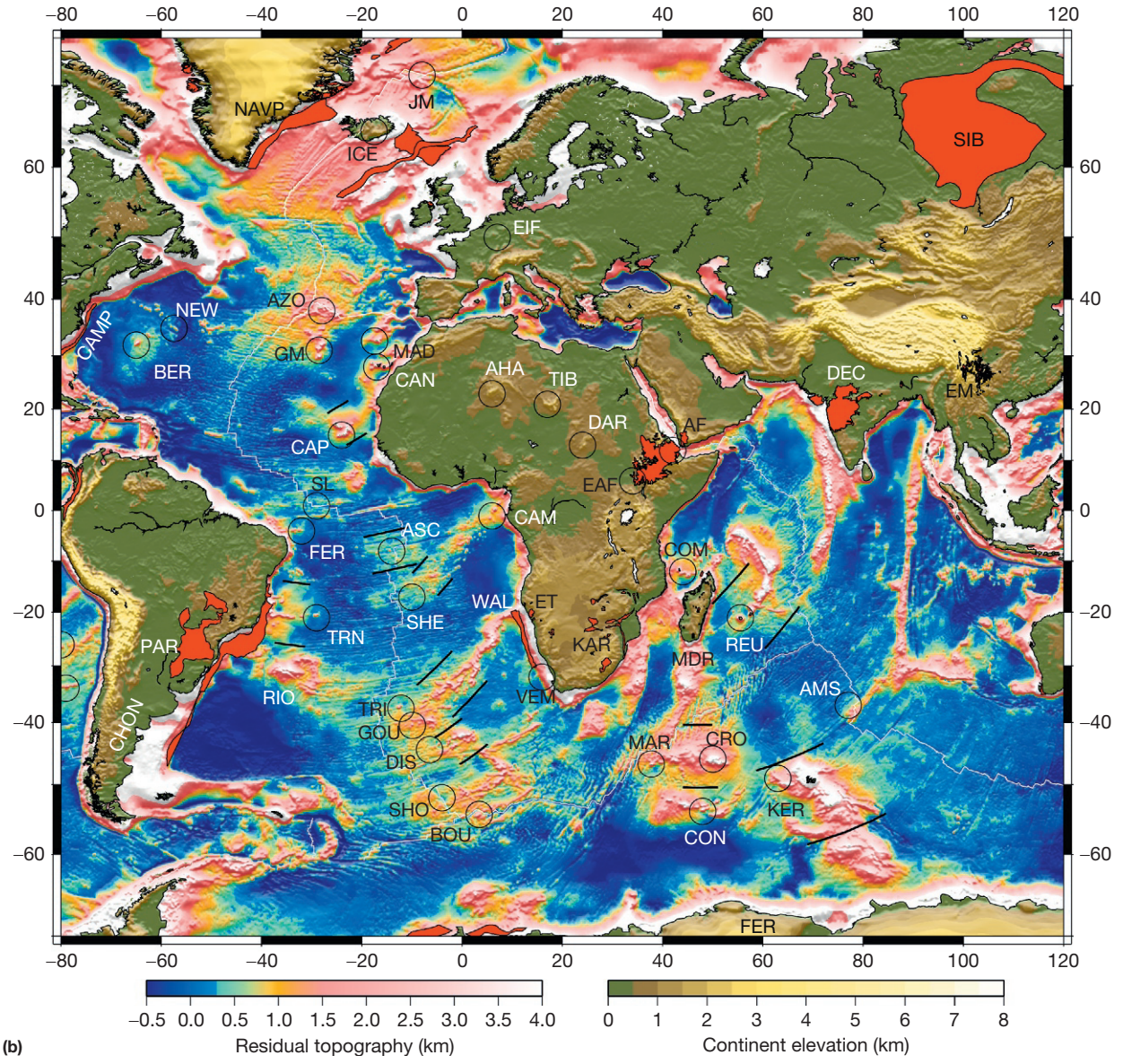


Figure 1 Elevation on the continents and residual topography in the oceans. Residual topography is the predicted bathymetry grid of Smith and Sandwell (1997) corrected for sediment loading and thicknesses (Laske and Masters, 1997) and for seafloor subsidence with age (Stein and Stein, 1992) using seafloor ages from Müller et al. (1993a) (areas without ages are interpolated using cubic splines). Grid processing and display was done using GMT (Wessel and Smith, 1995). Color change from blue to turquoise is at 300 m and delineates the approximate boundaries of anomalously shallow seafloor. Circles mark estimated locations of most recent (hotspot) volcanism. Pairs of lines are used to measure the widths of some of the hotspot swells (**Table 1**). Flood basalt provinces on the continents and continental margins are red. (a) Pacific region and (b) Indo-Atlantic region.

et al., 2004). The Canaries are unusual in that single volcanoes often remain active for tens of My (Geldmacher et al., 2005), much longer than, for example, the typical duration of the Hawaiian volcanoes' shield stage of ~1 My (e.g., Clague and Dalrymple, 1987; Ozawa et al., 2005). The long life span of some of the Canary volcanoes has contributed to uncertainty in defining a geographic age progression but is consistent with the slow motion of the African Plate over the hotspot.

Finally, Iceland is often cited as a classic Atlantic hotspot with volcanic tracks on two plates due to its location on a MOR (**Figure 3**). The thickest crust of the hotspot tracks occur along the Greenland–Iceland and Faeroe–Iceland Ridges extending NW and SE from Iceland. Anomalously thick oceanic crust immediately adjacent to these ridges shows datable magnetic lineations (Jones et al., 2002b; Macnab et al., 1995; White, 1997). Extrapolating the ages of these lineations onto the Greenland–Iceland and Faeroe–Iceland Ridges suggests the occurrence of age-progressive volcanism with seafloor spreading, which is most easily explained by the Iceland hotspot causing excess magmatism very near to or at the Mid-Atlantic Ridge since the time of continental breakup (White, 1988, 1997; Wilson, 1973). Earlier volcanism occurring as flood

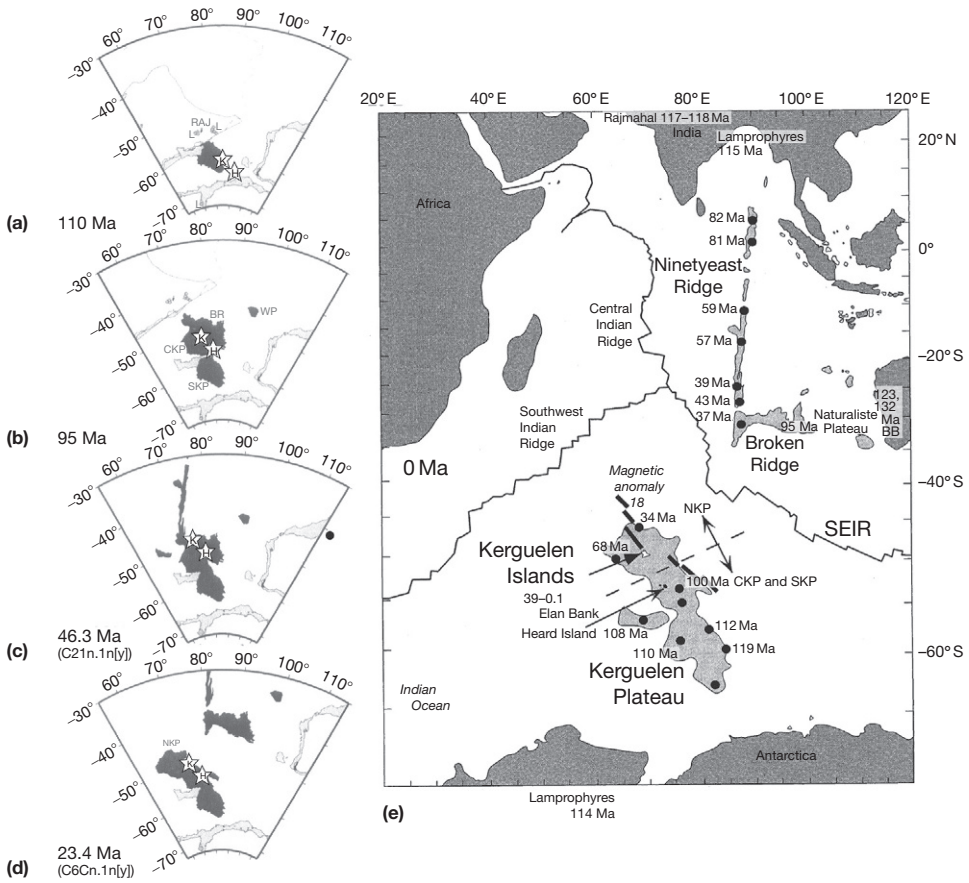


Figure 2 Evolution of the Kerguelen Plateau and hotspot track reproduced from (a) to (d) Coffin MF, Pringle MS, Duncan RA, et al. (2002) Kerguelen hotspot magma output since 130 Ma. *Journal of Petrology* 43: 1121–1139. (a) The initial pulse of volcanism formed the Rajmahal Traps (RAJ), lamprophyres (L) and the Southern Kerguelen Plateau (SKP) from ~120 to ~110 Ma. Stars mark reconstructed positions of the hotspot (Müller et al., 1993b) assuming a location at Kerguelen archipelago (K) and Heard Island (H). By 110 Ma, seafloor spreading between India, Antarctica, and Australia is well under way. The formation of the RAJ by the Kerguelen hotspot requires the hotspot to have moved by ~10° relative to the Earth's spin axis (Kent et al., 2002; cf. **Figure 9**). (b) By 95 Ma, central Kerguelen Plateau (CKP) and Broken Ridge (BR) have formed. (c) At 46.3 Ma, northward migration of the Indian Plate formed the Ninetyeast Ridge along a transform fault in the MOR system. (d) The Southeast Indian Ridge has propagated northwestward to separate CKP and BR (by ~42 Ma), which continue drifting apart at 23.4 Ma. (e) Current configuration as shown by Nicolaysen et al. (2000) but with ages from Coffin et al. (2002). Bold line along NE margin of the plateau marks magnetic anomaly 18 (41.3–42.7 Ma).

basalts along the continental margins of Greenland, the British Isles, Norway, and Baffin Island at about 56–62 Ma marks a conservative estimate for the age of the Icelandic melting anomaly (e.g., Lawver and Mueller, 1994; Saunders et al., 1997; White and McKenzie, 1989).

At a few volcano chains, the continuous record of volcanism is relatively short (a few tens of My) and the full duration of volcanism is poorly constrained. For example, the New England Seamount chain records ~20 My of oceanic volcanism (Duncan, 1984), but an extrapolation to the volcanic provinces in New England could extend the duration by another 20 My (cf. O'Neill et al., 2005). The duration of the Azores hotspot is also unclear. Gente et al. (2003) hypothesized the Azores hotspot and the Great Meteor and Corner Rise Seamounts (~85 Ma) to be conjugate features. Yet, age constraints of these features are poor and any geochemical association with the Azores group remains unknown.

In the Pacific Ocean (see Clouard and Bonneville (2005) and references therein), the total durations of the Cobb (1.5–29.2 Ma; Desonie and Duncan, 1990; Turner et al., 1980) and Bowie–Kodiak chains (0.2–23.8 Ma; Turner et al., 1980) remain unknown, as they terminate at the subduction zone south of Alaska. The Easter chain extends to at least ~30 Ma on the Nazca Plate (Clouard and Bonneville, 2005; Ray et al., 2012). On the Pacific Plate, the record extends perhaps even further into the past, but the tentative link with the Tuamotu Plateau (Clouard and Bonneville, 2001; Ito et al., 1995) clearly requires further dating.

7.10.2.1.2 Short-lived age-progressive volcanism

At least ten volcanic provinces show age-progressive volcanism lasting less than 25 My (e.g., Clouard and Bonneville, 2005). The Samoan hotspot has been active for 20 My (Hart et al., 2004) and possibly longer (Duncan and Clague, 1985). The hotspot-like age progression along the Samoan track is obscured by voluminous secondary volcanism covering the main islands of the archipelago (e.g., Natland, 1980) but has been revealed by recent sampling of the submarine flanks of the volcanoes (Koppers et al., 2008, 2011b). The Society and Marquesas Islands represent voluminous volcanism but for

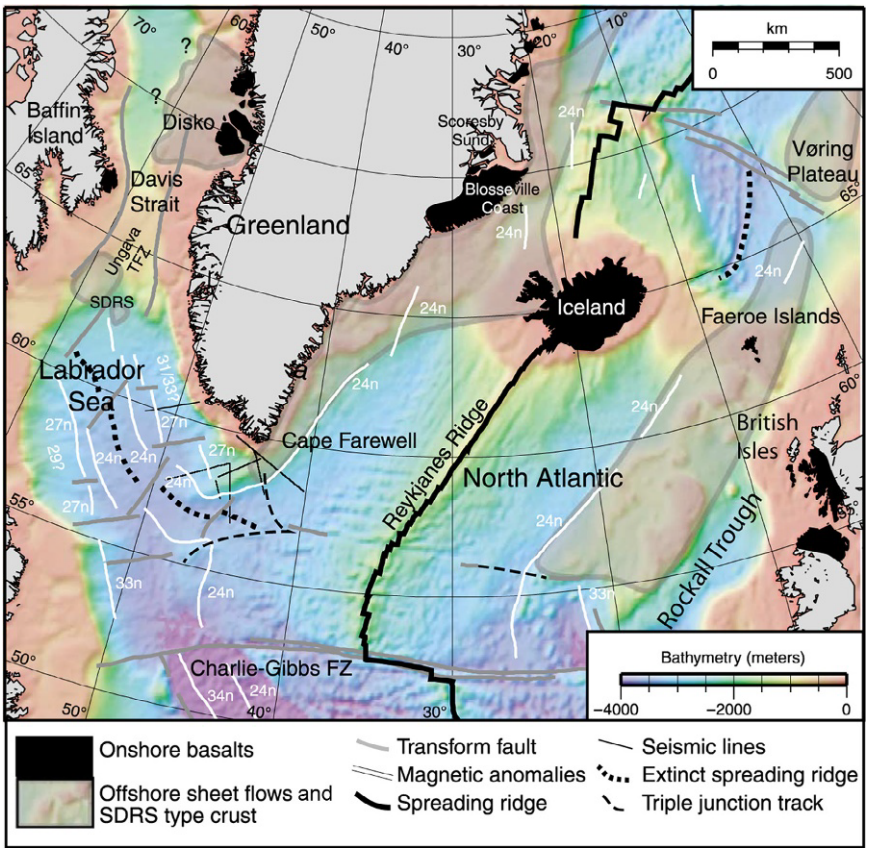


Figure 3 From Nielsen et al. (2002). Distribution of Paleogene flood basalts in the North Atlantic Volcanic Province. Selected seafloor magnetic lineations, major transform faults, active and extinct spreading ridges, and seismic lines are marked as labeled.

geologically brief durations of ~4.6 My and ~4.7 My, respectively (Chauvel et al., 2012; Guillou et al., 2005; Legendre et al., 2006). Clouard and Bonneville (2001) argued that the Marquesas hotspot could have formed the Line Islands and Hess Rise based on geometric arguments. If this interpretation is correct, the large gap in volcanism between the three provinces implies strongly episodic volcanism. However, the age trend of the Marquesas volcanoes departs from Pacific Plate motion in both speed and direction and thus contradicts a hotspot origin altogether (Legendre et al., 2006). Similar to the Marquesas, the Foundation chain (duration from 2.1 to 21 Ma) displays an age progression that is somewhat faster than Pacific Plate motion (O'Connor et al., 2004), which could suggest an absolute drift of the underlying melting anomaly (cf. O'Neill et al., 2005). Short hotspot durations of 11–24 My further occur along the Pitcairn (0–11 Ma), Caroline (1.4–13.9 Ma), Tarava (35.9–43.5 Ma; Clouard et al., 2003), Tokelau (58–74 Ma; Koppers et al., 2007), and Japanese seamount (94–118 Ma; Koppers et al., 2003) chains on the Pacific Plate, as well as the Tasmantid seamounts near Australia (7–24.3 Ma; McDougall and Duncan, 1988; Müller et al., 1993b). Such short hotspot durations do not fit in with the classical explanation that hotspots are underlain by long-lived mantle plumes.

An intriguing yet enigmatic form of age-progressive volcanism is represented by the Pukapuka, Sojourn, and Hotu Matua ridges, which trend WNW–ESE and away from the East Pacific Rise. With respect to its geographic trend and duration, the Pukapuka ridge resembles some of the other volcano chains in the region, such as the Foundation chain (Maia et al., 2001; O'Connor et al., 2002, 2004), but it stands out because of its age pattern as well as its smaller volcanic volume and morphology: the Pukapuka is not built of discrete conical seamounts, but rather of en echelon elongate ridges (Lynch, 1999). Moreover, Pukapuka's rapid and variable rate of age progression (Janney et al., 2000; Sandwell et al., 1995) suggests that the magma source has migrated eastward by rates of ~20 cm year^{-1} (**Figure 4**), which is difficult to reconcile with the concept of a near-stationary hotspot source. While the Sojourn and Hotu Matua ridges lack the lateral extent to robustly define an age progression, they appear very similar to the parallel Pukapuka ridge, not only in terms of the available ages and the morphology of the volcanic edifices (Forsyth et al., 2006) but also in terms of their association with gravity, bathymetry, and seismic wave speed anomalies (**Figure 4**; Harmon et al., 2011).

7.10.2.1.3 Non-age-progressive volcanism

An important form of oceanic volcanism does not involve simple geographic age progressions. Amsterdam–St. Paul and Cape Verde are two examples that each represent opposite extremes in terms of size and duration. The Amsterdam–St. Paul hotspot represents a relatively small and short-lived (≤5 My) melting anomaly on the Southeast Indian Ridge that

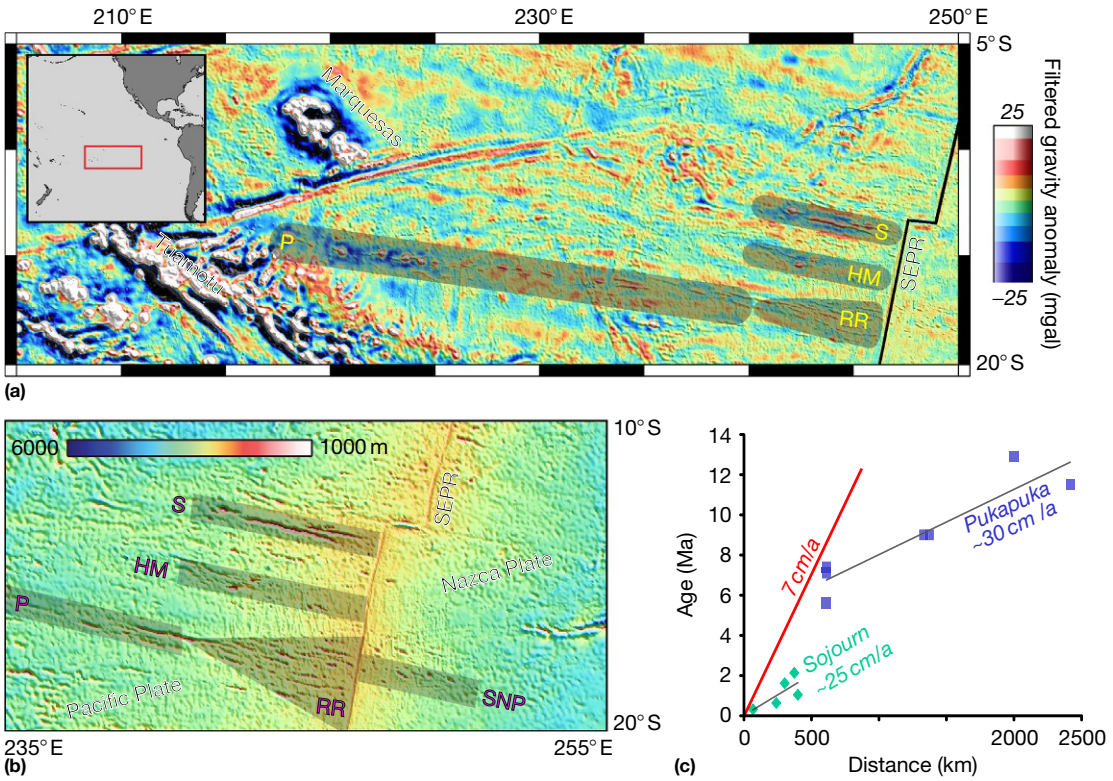


Figure 4 From Ballmer et al. (2013a). (a) Filtered gravity anomalies (for filtering scheme, see Harmon et al., 2011) highlighting the east–west trending 'Haxby' gravity lineations (Haxby and Weissel, 1986) and (b) bathymetry in the southeast Pacific (SEPR: southern East Pacific Rise). Volcanic ridges on the Pacific Plate (S: Sojourn; HM: Hotu Matua; and P: Pukapuka), the Rano Rahi seamount field (RR), and a short seamount chain on the Nazca Plate (SNP) are shaded. (c) Ages collected along the Pukapuka and Sojourn ridges scatter about trends that are much faster (by ~20 cm year^{-1}) than Pacific Plate motion (red line). Ballmer et al. (2013a) argued that the Pukapuka source crossed the SEPR at ~5 Ma to form the apparently young (≤3 Ma) SNP.

is geochemically separate from the Kerguelen hotspot (cf. Doucet et al., 2004; Graham et al., 1999) Its small size and duration, as well as its location on a MOR, likely contribute to the lack of an identifiable age progression. The Cape Verde volcanoes, on the other hand, are larger and have likely existed since Early Neogene (e.g., McNutt, 1988). The islands also lack a clear linear age progression, but this can be expected for a hotspot located very close to the Euler pole of the African Plate (McNutt, 1988).

Examples of volcano chains with complex age–space relations on the rapidly moving Pacific Plate are more difficult to explain within the hotspot framework. While the lack of modern dates on samples of the Mid-Pacific Mountains and the Geologist seamounts precludes robust interpretations, the Marcus–Wake seamounts and Gilbert Ridge, as well as the Marshall, Magellan, and Line Islands, clearly display synchronous or near-synchronous volcanism over large distances of 1500–2000 km (e.g., Davis et al., 2002; Koppers et al., 2003; Clouard and Bonneville, 2005 and references therein). These volcano groups also show long-lived, recurrent activity at individual islands (**Figure 5**). They can thus more reasonably be explained by 'hot lines' or melting anomalies that are elongated in the direction of plate motion (Bonatti and Harrison, 1976; Bonatti et al., 1977) or even by much broader magmatic anomalies with melt extraction controlled by the lithosphere (see **Section 7.10.3.7.3**). The Cameroon volcanic line, which extends from Cameroon to the SW straddling the passive continental margin of Africa (**Figure 6**), may represent an analogue to the Pacific cases of volcanism with coeval activity over distances greater than 1000 km (Marzoli et al., 2000; Montigny et al., 2004).

Recent volcanism along the prominent Cook–Austral Islands in the South Pacific also displays overall complex geographic age patterns (McNutt et al., 1997) and complex uplift history at individual volcanoes (Adam and Bonneville, 2008; Jordahl et al., 2004). The geographic age pattern of a younger episode of volcanism (0–20 Ma) alone requires either the activity of two parallel hot lines (Ballmer et al., 2009) or that of four closely spaced, short-lived hotspots (representing the Macdonald, Arago/Rurutu, Raivavae and Rarotonga tracks; Bonneville et al., 2002; 2006) (cf. dashed lines in **Figure 5 (c)**). In the framework of a hotspot interpretation, geochemical affinities with seamounts near the Samoan Islands have been suggested to indicate long life spans (of up to ~25 and ~50 My) for some of the tracks (Arago/Rurutu and Macdonald, respectively) (Jackson et al., 2010).

The remaining oceanic hotspots in **Table 1** do not have sufficient data to define time–space relations. This list includes volcano chains that simply have not yet been adequately dated (Socorro, Tuamotu, and Great Meteor), hotspots that have small footprints (e.g., Discovery, Ascension/Circe, Sierra Leone, Conrad, Bermuda), or spatial patterns

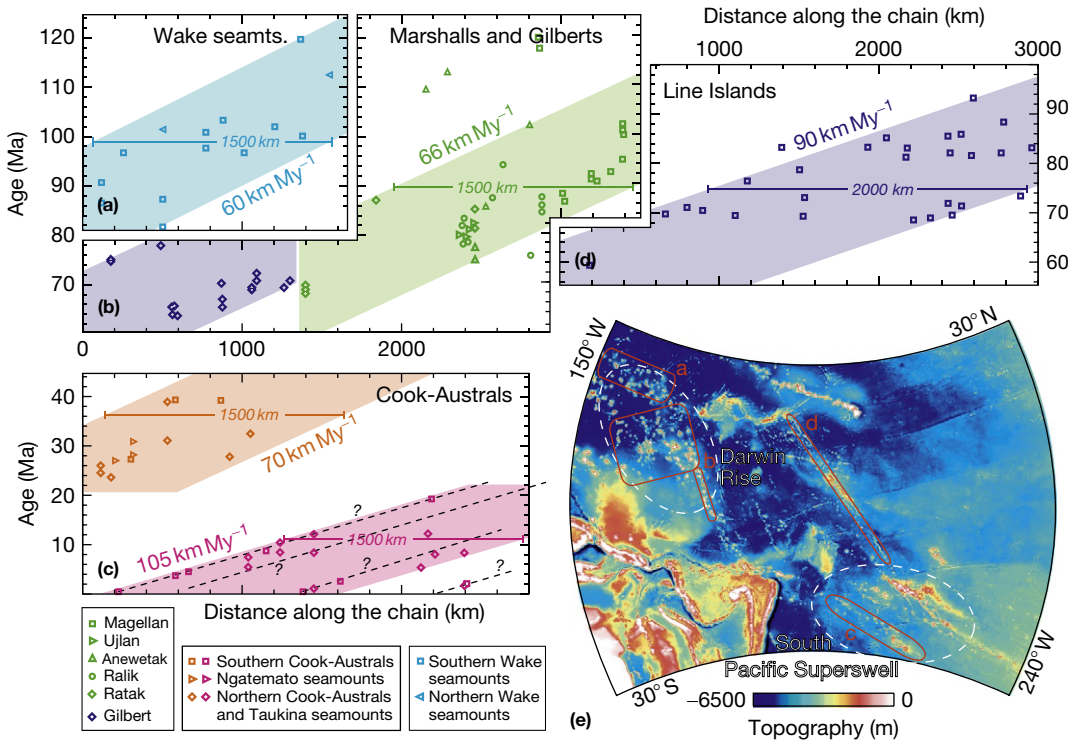


Figure 5 Compilation of volcano ages sampled along the (a) Wake seamounts, (b) Marshall Islands and Gilbert Ridge, (c) Cook–Austral Islands, and (d) the Line Islands (see (e) for locations). For (a–c), sample ages collected along each of the subparallel trends (different symbols) are geometrically projected onto the same distance axis. Dashed lines in (c) indicate the putative Macdonald, Raivavae, Arago/Rurutu, and Rarotonga (from left to right) hotspot tracks as identified by Bonneville et al. (2002). As shown by the shaded fields, the scattered volcano ages are alternatively explained by the activity of hot lines of lengths ~1500 km (a–c) or up to ~2000 km (d) for the Line Islands. Modified from Ballmer MD, van Hunen J, Ito G, Bianco TA, and Tackley PJ (2009) Intraplate volcanism with complex age-distance patterns: A case for small-scale sublithospheric convection. *Geochemistry, Geophysics, Geosystems* 10.

controlled by interaction with spreading centers (e.g., Bouvet and Balleny).

7.10.2.1.4 Continental hotspots and melting anomalies

There are a number of continental melting anomalies that are not directly associated with present-day subduction, orogeny, or continental rifting. Such activity occurs, for example, in Mongolia (Hunt et al., 2012), Antarctica (e.g., Giordano et al., 2012), western North America (Lipman and Glazner, 1991; Nelson and Tingey, 1997), eastern Australia (e.g., Griffin and McDougall, 1975), eastern China, Europe, and Africa (Lustrino and Wilson, 2007; green dots in Figure 7). A number of these areas are also shown in Figure 1 and listed in Table 1, such as the North American Yellowstone (YEL)–Snake River Plain volcanic progression (YSRP), the European Eifel hotspot (EIF) (cf. Lustrino and Wilson, 2007 and references therein), and the African hotspots such as expressed by the volcanic mountains of Jebel Mara (DAR), Tibesti (TIB), and Ahaggar (HOG) (Burke, 1996).

The Yellowstone hotspot stands out among these, because it is the only one that displays a clear age progression. It is recorded by silicic caldera forming events starting at 16–17 Ma at the Oregon–Nevada border at a distance of ~700 km from the hotspot. Initial rhyolitic volcanism is followed by long-lived basaltic volcanism along the present-day Snake River Plain (SRP) (Figure 8). The effective speed of the hotspot track is 4.5 cm year^{-1}, which includes a component of the present-day plate motion (2.5 cm year^{-1}) and a component caused by Basin and Range extension. The YSRP track appears to be mirrored by the westward-progressing High Lava Plains (HLP) track in southern Oregon (Kincaid et al., 2013; Long et al., 2012; Figure 8). The lack of age progression of the other hotspots is usually attributed to the nearly motionless African Plate and the slow motion of the European Plate.

7.10.2.1.5 The hotspot reference frame

The existence of many long-lived volcano chains with clear age progressions led Morgan (1971, 1972) to suggest that hotspots remain relatively stationary to one another and therefore define a global kinematic reference frame separate from plate motions (Duncan and Clague, 1985; Morgan, 1983). Such an absolute reference frame was used to anchor the well-constrained relative motions of the tectonic plates with respect to the lower mantle and to quantify absolute displacements of the Earth's spin axis (i.e., true polar wander) through geologic time (Biggin et al., 2012; Torsvik et al., 2002, 2010b, 2012). The concept of a fixed hotspot reference frame, however, has been challenged by observations that indicate relative motions between hotspots, particularly between the Indo-Atlantic and Pacific hotspot groups (Acton and Gordon, 1994; DiVenere and Kent, 1999; Molnar and Stock, 1987; Raymond et al., 2000; Tarduno and Gee, 1995; Torsvik et al., 2002). Relative motion between these two groups is slow but robust and independent of the plate circuit chosen to

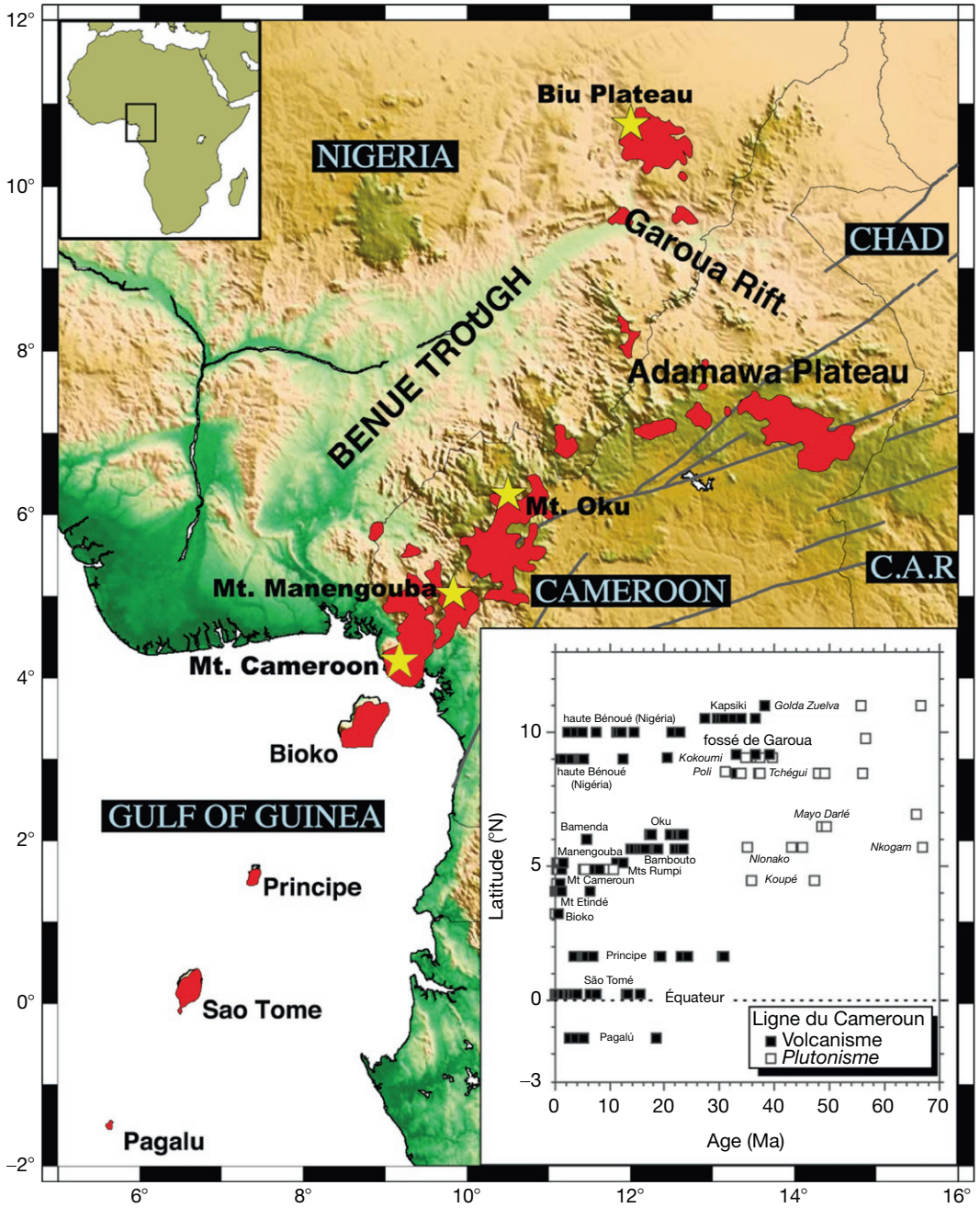


Figure 6 Modified from Gallacher RJ and Bastow ID (2012) The development of magmatism along the Cameroon Volcanic Line: Evidence from teleseismic receiver functions. *Tectonics* 31. Sketch map of the Cameroon volcanic line with magmatic provinces shown red. Yellow stars denote selected volcanoes as annotated. Inset (reproduced from Montigny R, Ngounouno I, and Deruelle B (2004) K-Ar ages of magmatic rocks from the Garoua rift: Their place in the frame of the 'Cameroun Line'. *Comptes Rendus Geoscience* 336: 1463–1471) shows age compilation of plutonic and volcanic rocks along the Cameroon Line.

connect the Indo-Atlantic and Pacific hemispheres. Accordingly, uncertainties in the hotspot reference frame inevitably increase looking back into the geologic past.

In addition to slow motion between the Pacific and the Indo-Atlantic hotspot groups, somewhat slower relative motion occurs among hotspots within each of the two groups. In the Pacific, relative motion has been inferred for many hotspot pairs but is best documented from interchain distances of coeval volcanoes within the well-sampled Hawaii and Louisville chains (Koppers and Staudigel, 2005; Koppers et al., 1998, 2011a, 2012; Wessel and Kroenke, 2008, 2009). Most importantly, the rapid southward motion of the Hawaiian hotspot while forming the Emperor seamounts before ~50 Ma has been identified to modulate these distances (Pares and Moore, 2005; Tarduno et al., 2003, 2009). In the Indian Ocean, the inferred absolute southward motion of the

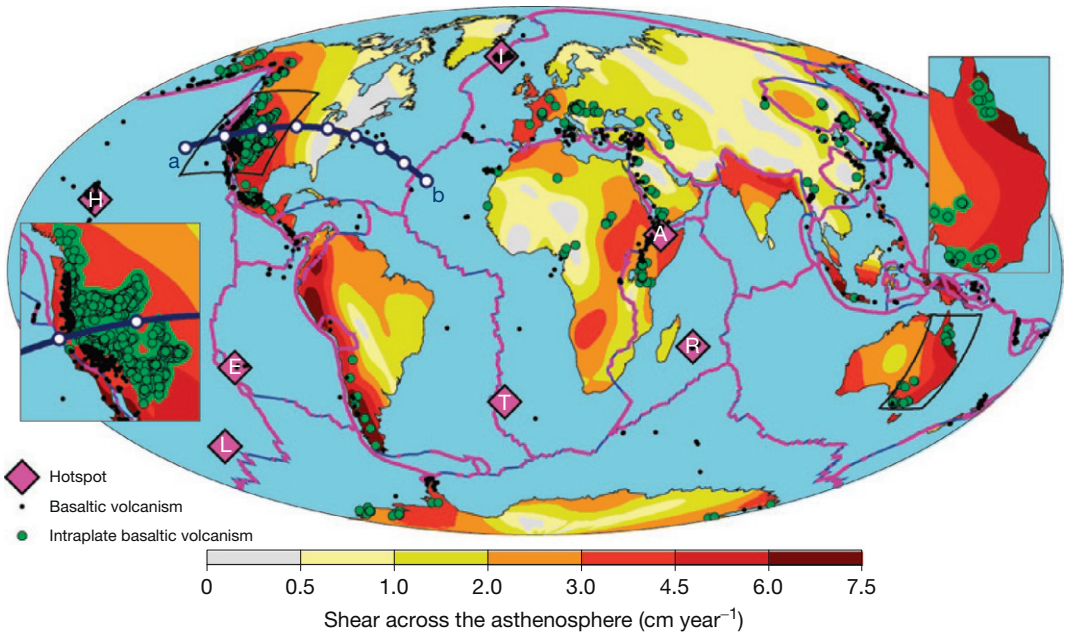


Figure 7 From Conrad et al. (2011). Shear across the asthenosphere beneath (colors), and basaltic intraplate volcanism on (green circles) continents. Asthenospheric shear is shown as the amplitude of the vector difference between the motions of the plates and the underlying mantle (i.e., at the base of the asthenosphere as predicted by Conrad and Behn, 2010). Primary (putatively plume-fed) hotspots (Courtillot et al., 2003) are shown as purple squares. Excluding volcanism that is close to plate boundaries or primary hotspots (small black dots), continental intraplate basaltic volcanism (green circles) occurs more frequent where shear across the asthenosphere is large. Reproduced from Conrad CP, Bianco TA, Smith EI, and Wessel P (2011) Patterns of intraplate volcanism controlled by asthenospheric shear. *Nature Geosciences* 4: 317–321.

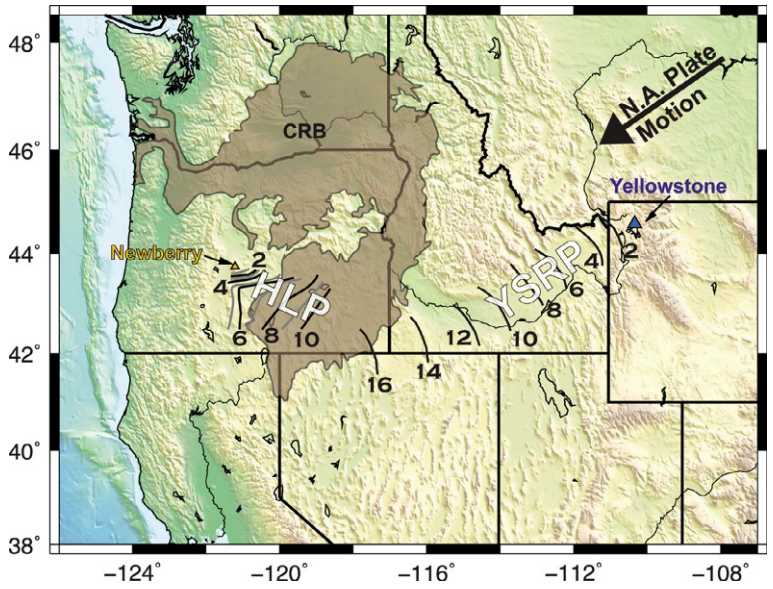


Figure 8 Topographic map with age progressions along the Yellowstone–Snake River Plain (YSRP) and High Lava Plains (HLP) tracks. Ages of volcanism are annotated (triangles denote centers of active volcanism). Columbia River basalts (CRBs; age: 16.6–15 Ma) are shown as brown shading. Modified from Long MD, Till CB, Druken KA, et al. (2012) Mantle dynamics beneath the Pacific Northwest and the generation of voluminous back-arc volcanism. *Geochemistry, Geophysics, Geosystems* 13.

Kerguelen hotspot since 100 Ma hinges on the weakly constrained location of its present location (Steinberger and Antretter, 2006; Weis et al., 2002). At or near present day (i.e., <4–7 Ma), motion between hotspots globally is unresolvable or at least much slower than it has been in the geologic past (Gripp and Gordon, 2002; Wang and Wang, 2001). For a more complete description of these issues, we refer the reader to **Chapter 6.02**.

Whereas fixity of the source of volcanism relative to the drifting plates has been widely used as diagnostic test for a

hotspot origin, slow relative motion of hotspots is naturally expected in the framework of mantle plume theory. Rising plumes are disturbed by large-scale mantle convection that should slowly yet steadily displace their surface expressions. Within this concept, the drift between hotspots is limited due to the relatively sluggish convection in the lower mantle combined with the potential for chemical anchoring of plumes at the base of the mantle (cf. Davaille et al., 2002; Jellinek and Manga, 2002; McNamara and Zhong, 2004; see also **Chapter 7.11**). Global geodynamic models that simulate plumes rising in a convecting mantle predict directions and amplitudes of motion between hotspots that are consistent with observations for a large range of model parameters (Steinberger, 2000; Steinberger and Antretter, 2006). For example, these models can reproduce the westward motion of the Iceland hotspot relative to both the Indo-Atlantic and Pacific hotspots (Norton, 2000; Raymond et al., 2000) for a relatively shallow plume source near the transition zone (Mihalffy et al., 2008). Also, they routinely predict independent drift for a given hotspot pair (**Figure 9**), as is evident for the Hawaii and Louisville hotspots (Koppers et al., 2012). Such model predictions have been used to define 'moving hotspot' reference frames (Doubrovine et al., 2012). However, these reference frames yield similar results as the single fixed hotspot reference frame in explaining volcano age–distance relations when considering the associated uncertainties (O'Neill et al., 2005). Future work is needed to be able to use hotspot motions as a constraint, for example, to infer mantle properties from the best-fitting geodynamic models.

Sudden changes in the direction of hotspot drift are routinely predicted by many of these models (Steinberger and Antretter, 2006; Steinberger et al., 2004; **Figure 9**), and these can further contribute to bends in hotspot chains, with the Hawaiian–Emperor bend near 50 Ma being a prominent example. The classical explanation for such bends invoked sudden changes in plate motion over fixed hotspots (Andrews et al., 2006; Sharp and Clague, 2006; Whittaker et al., 2007). As a general explanation, however, it is challenged by evidence for prominent bends in seamount chains on the same plate that did not occur at the same time (Koppers et al., 2007). In addition, paleomagnetic evidence along the Hawaiian–Emperor chain indicates a sudden change in hotspot motion from being dominantly southward to being nearly stationary (in the N–S sense) right near the 50 My bend (Kono, 1980; Sager et al., 2005; Tarduno et al., 2009). A combination of changes in both hotspot and plate motion may therefore have occurred during the formation of the Hawaiian–Emperor bend (O'Connor et al., 2013).

7.10.2.2 Topographic Swells

Figure 1 illustrates that most oceanic hotspots and melting anomalies are associated with anomalously shallow topography that extends several hundred km beyond the area of excess volcanism. For individual hotspots, we identify the presence of such a seafloor swell if residual topography exceeds an arbitrarily chosen value of 300 m and extends appreciably (>100 km) away from the center of volcanism. Such hotspot

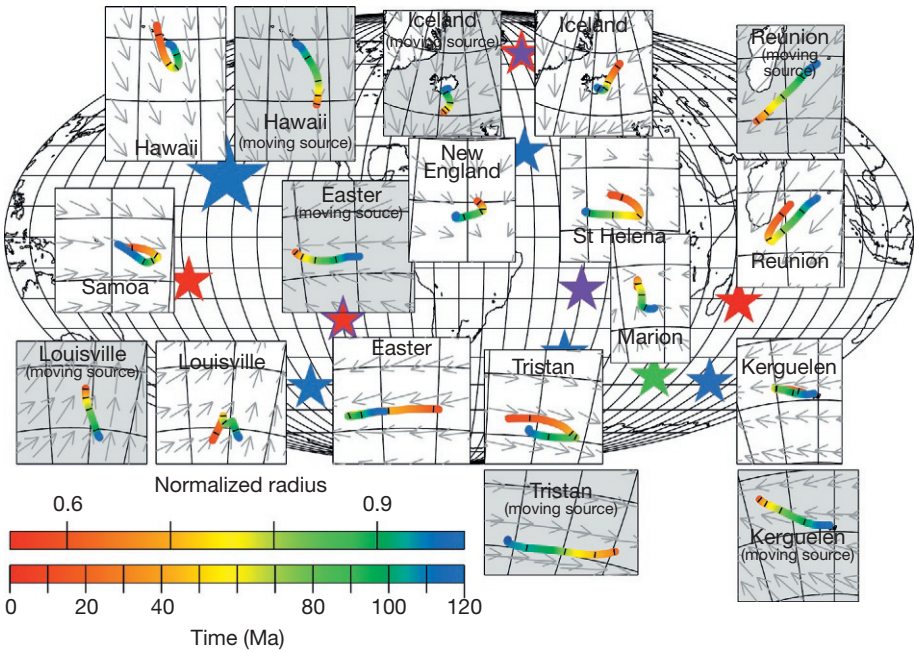


Figure 9 Hotspot motions predicted from a global convection model. Hotspot positions today are shown as stars and through time as rainbow-colored bands in insets (scale at the very bottom). Star fill color reflects normalized radius of the plume conduit (top scale). For all hotspots shown, white insets display the predicted hotspot motions for a fixed plume source (i.e., anchored plume). For a subset of hotspots, gray insets display the predicted hotspot motions for a moving plume source. Also shown are the horizontal flow components in the lower mantle (at either 2512 km (gray boxes) or 1918 km depth (white boxes)). Arrow length scale is 5° of arc per cm/year. Reproduced and modified from Steinberger B and Antretter M (2006) Conduit diameter and buoyant rising speed of mantle plumes: Implications for the motion of hotspots and shape of plume conduits. *Geochemistry, Geophysics, Geosystems* 7.

swells are indeed very common. Near major hotspots, they rise to ~2 km above the surrounding seafloor and extend many hundreds of km away from the center of volcanic activity (Crough, 1983). They are also apparent on chains with very small volcanoes such as the Socorro and Tasmantid tracks, as well as the Pukapuka and Sojourn ridges (Harmon et al., 2006, 2011). In contrast, the Madeira and Canary hotspots are two cases that break this pattern. The lack of obvious swells around these large and long-lived hotspots indeed motivates more detailed investigation.

Hotspot swells are typically restricted to the younger portions of the volcano chain (**Table 1**). The Hawaiian swell, for example, is prominent beneath the youngest (southeast) part of the Hawaiian ridge but begins to fade near ~178°W and disappears near the Hawaiian–Emperor bend (~50 Ma). Swells also decay along the Louisville (near a volcano age of ~34 Ma) and the Tristan chains (on the Walvis Ridge near 62–79 Ma). Similar behavior is seen for the Kerguelen track with shallow seafloor that extends away from the Kerguelen Plateau on the Antarctic Plate juxtaposed to normal topography around the southernmost portion of Broken Ridge (~43 Ma) on the Indian Plate. Swells associated with hotspot tracks in the South Atlantic also systematically decay with age, perhaps except for that associated with the St. Helena track, which however is juxtaposed to the active Cameroon volcanic line. Old chains that lack swells in our analysis include the Japanese–Wake seamounts (>70 Ma), the Magellans (>70 Ma), the Mid-Pacific Mountains (>80 Ma), and the Musicians (>65 Ma). This suggests that if a swell forms at a hotspot, it decreases in height with time until it can no longer be detected at an age of ~50 Ma or less.

Whereas swells are more difficult to measure on the thick and compositionally heterogeneous continental plates, the Yellowstone hotspot is associated with well-documented excess topography. The measured topographic bulge is 600 m high and ~600 km wide centered around the Yellowstone caldera (cf. **Figure 8**) and is associated with high heat flow, extensive hydrothermal activity, and a 10–12 m positive geoid anomaly (Smith and Braile, 1994). Systematic topographic decay along the track in the SRP is consistent with thermal contraction in response to the progression of the American Plate over a hotspot (Smith and Braile, 1994).

The South Pacific Superswell is a much larger swell that spans ~3000 km and supports the hotspots in French Polynesia (Adam and Bonneville, 2005; Hillier and Watts, 2004; McNutt and Fischer, 1987). Other swells of comparable size include the ancient Darwin Rise in the northwestern Pacific (McNutt et al., 1990) and the African Superswell (e.g., Nyblade and Robinson, 1994), which encompasses the southern portion of Africa and the South Atlantic down to the Bouvet Triple junction. Both the South Pacific and African Superswells are associated with active volcanism, geoid anomalies (Adam et al., 2010; Cadio et al., 2011), and broad seismically slow structures in the lower mantle (e.g., Dziewonski et al., 2010; Lekic et al., 2012; see also **Section 7.10.2.5.1** in the succeeding text). While the oceanic hotspots near Africa to first order display the key characteristics predicted by hotspot theory (e.g., O'Connor et al., 2012), the hotspots in the South Pacific are mostly short-lived, including some with extreme geochemical signatures (Konter et al., 2008; Koppers et al., 2003; Staudigel et al., 1991) and complex age progressions (see **Section 7.10.2.1**; Ballmer et al., 2009; McNutt, 1998; McNutt et al., 1997). The Darwin Rise is often interpreted as the Cretaceous counterpart of the South Pacific Superswell (Cadio et al., 2011; Koppers et al., 2003; McNutt and Judge, 1990). We will discuss the geodynamic mechanisms for swell and superswell support in **Sections 7.10.3.4** and **7.10.3.7.1**, respectively.

7.10.2.3 Flood Basalt Volcanism

Large igneous provinces provide further constraints on the nature of hotspots and mantle dynamics. LIPs are characterized by massive outpourings and intrusions of basaltic lava, often occurring within a couple of million years. Fundamental reviews of the nature and possible origin of LIPs are provided by Richards et al. (1989), Coffin and Eldholm (1994), Mahoney and Coffin (1997), Courtillot and Renne (2003), and Bryan et al. (2010). As characterized by Coffin and Eldholm (1994), intraplate continental LIPs (such as the Siberian Traps or the Columbia River basalts) often form by fissure eruptions and horizontal flows of massive tholeiitic basalts. Another form of LIP volcanism occurs at passive margins (such as the North Atlantic Volcanic Province or Etendeka–Paraná) that are associated with continental rifting (White and McKenzie, 1989). The initial prerift pulse of volcanism tends to be rapid and often followed by a more prolonged production of thicker-than-normal oceanic crust and sometimes by long-lived hotspot activity. Oceanic LIPs form broad, flat-topped features of anomalously thick crust with some eruptions being subaerial (Kerguelen oceanic plateau). Other eruptions appear to be confined to below sea level (e.g., Ontong Java Plateau and Shatsky Rise). Regional uplift with often rapid onsets commonly accompanies the formation of continental and oceanic LIPs (Saunders et al., 2007). This section reviews some basic geophysical and geologic observations of LIPs that formed since the Permian. For information on older LIPs, we refer the reader to White and McKenzie (1995), Ernst and Buchan (2001; 2003), and Ernst (2007). **Figure 1** shows locations with indicated abbreviations; **Figure 10** summarizes the areas and volumes of the provinces described in the succeeding text.

7.10.2.3.1 Intraplate continental LIPs

Columbia River Basalts (*CRBs*). The Columbia River basalts erupted a volume of ~0.17 Mkm between 16.6 and 15.3 Ma (Courtillot and Renne, 2003). Excellent exposures provide insights into flow structures and relationships to feeder dikes. Individual eruptions have volumes in excess of 2000 km^3 and flow over distances up to 600 km (Hooper, 1997). The lack of collapse structures suggests that large amounts of magma were rapidly derived from the base of the crust (Hooper, 1997). The volumetrically most important pulse of volcanism is the Grande Ronde basalt that was emplaced in less than 0.5 My (Barry et al., 2010). Detailed geochemical work suggests that the CRBs rose from a centralized magmatic system following crustal assimilation of plume-derived magmas (Wolff et al., 2008).

Emeishan (*EM*). The Emeishan province in western China is estimated to have spanned an area of at least 2 Mkm^2 and volume of 1 Mkm^3 when it first formed (Zhou et al., 2002).

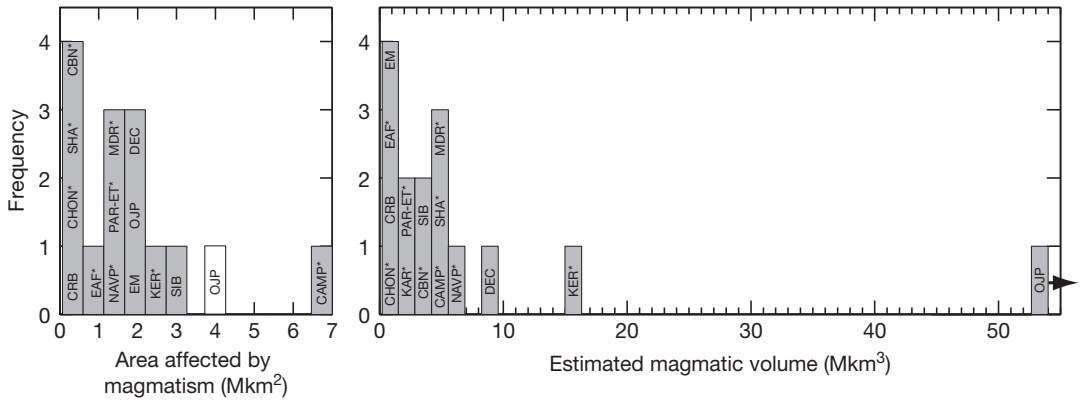


Figure 10 Histograms of estimated areas and volumes of the LIPs discussed in the text. The OJP–Manihiki–Hikurangi LIP would have the area shown by white bar and a larger volume than shown, as indicated by arrow.

Eruption ages have been measured at ~258 Ma (Courtillot and Renne, 2003), which is confirmed by a zircon U–Pb date of ~259 Ma (Ali et al., 2005) and by more precise recent U–Pb ages (Shellnutt et al., 2012). Granitic intrusions indicate significant crustal assimilation in magmas that were sourced from the mantle (Shellnutt and Jahn, 2011; Xu et al., 2007; Zhong et al., 2007) A rapid, kilometer-scale uplift preceded the basaltic eruptions by 3 My (He et al., 2003; Sun et al., 2010; Xu et al., 2004). Basalts erupted rapidly (Zheng et al., 2010) and were accompanied by high-MgO basalts (He et al., 2003) and contemporaneous felsic and mafic intrusions (Zhong et al., 2011). Since their eruption, the Emeishan Traps have been fragmented and eroded, currently encompassing an area of ~0.3 Mkm^2 (Xu et al., 2001).

Siberia (*SIB*). The Siberian Traps occupy at present only 0.4 Mkm^2 and have an average thickness of 1 km (Sharma, 1997). There are strong indications, however, that they extend below the sedimentary cover and into West Siberian basin (Reichow et al., 2005). Additional dikes and kimberlites suggest a maximum extent of 3–4 Mkm^2 with possible extrusive volume of >3 Mkm^3. Individual flows can be as thick as 150 m and be traced over hundreds of kilometers. This large eruption took place within only 1 My and coincides with the Permo-Triassic boundary at 250 Ma (Courtillot and Renne, 2003; Reichow et al., 2009). The lack of significant sedimentary rocks or paleosols between flows confirms rapid extrusion (Sharma, 1997). The use of industry seismic and borehole data in the West Siberian basin shows that the basin elevation remained high during rifting, indicating dynamic mantle support (Saunders et al., 2005). Magnetic data elucidate that the eruption took place during oblique rifting (Allen et al., 2006), which may have increased the magmatic output. Petrologic and geodynamic models also agree that lithospheric thinning is required to reconcile the large magmatic output and lava geochemistry (Sobolev et al., 2011; White and McKenzie, 1995). Paleomagnetic reconstructions suggest that the Siberian Traps had erupted over the same part of the mantle where at ~60 Ma the North Atlantic Igneous Province formed (Smirnov and Tarduno, 2010), providing an interesting suggestion for a connection of both LIPs to a long-lived, deep, and nearly stationary source.

Yemen/Ethiopian/East African Rift System (*EAF*). An early volcanic episode in southernmost Ethiopia starting ~45 Ma was followed by widespread flood basalt volcanism in northwestern Ethiopia, Eritrea, and Yemen at ~30 Ma. The 30 Ma event consisted of tholeiites and ignimbrites (Pik et al., 1998) that erupted within 1–2 My (Ayalew et al., 2002; Hofmann et al., 1997). In Yemen, 0.35–1.2 Mkm^3 of mafic magmas were produced within 2 My, followed by less voluminous silicic volcanism starting ~29 Ma (Menzies et al., 1997). Flood volcanism appears to have occurred several My prior to the onset of extension along the East African Rift (EAR) at ~23 Ma (Hendrie et al., 1994; Morley et al., 1992) and in the Gulf of Aden at ~26 Ma (Menzies et al., 1997). Volcanism is bimodal with shield volcanoes forming on top of tholeiitic basalts (Courtillot and Renne, 2003; Kieffer et al., 2004). Currently, the Ethiopian and Kenyan rift systems are on an area of elevated topography of ~1000 km in diameter. A negative gravity anomaly suggests this topography is dynamically supported in the mantle (Stewart and Rogers, 1996). The analysis of primitive magmas indicates that the EAF magmas arise from mantle that is 140–170 K hotter than normal (Rooney et al., 2012).

Older continental flood basalts are often more difficult to detect in the geologic record due to the effects of surface uplift and erosion. For a review, see Ernst (2007). A general characteristic that is attributed to continental LIPs is radiating dike swarms (Mayborn and Lesher, 2004; Mege and Korme, 2004). These dike swarms provide the main pathways for basaltic magmas vertically from the mantle, as well as laterally over distances up to 2500 km (as suggested for the 1270 Ma Mackenzie dike swarm in N. America (Lecheminant and Heaman, 1989; Ernst and Baragar, 1992)).

7.10.2.3.2 LIPs near or on continental margins

Central Atlantic Magmatic Province (*CAMP*). The CAMP is primarily delineated by giant dike swarms and is associated with the early breakup of Gondwana between North Africa, eastern United States, and central South America. Ar/Ar dates suggest rapid production from 202 to 200 Ma over an area of 7 Mkm^2 and volume of >2 Mkm^3 (Courtillot and Renne, 2003; Hames et al., 2000; Marzoli et al., 1999), although more recent work puts the CAMP at slightly younger ages of ~199 Ma (Verati et al., 2007; Schaltegger et al., 2008; Jourdan et al., 2009; Marzoli et al., 2011).

Chon Aike (*CHON*). In contrast to the other LIPs discussed in the chapter, the Chon Aike province in Patagonia is primarily silicic with rhyolites dominating over mafic and intermediate lavas. The rhyolites may have formed due to intrusion of basalts into crust that was susceptible to melting. The province is relatively small with an area of 0.1 Mkm^2 and total volume of 0.235 Mkm^3 (Pankhurst et al., 1998). Chon Aike had an extended and punctuated eruptive history from Early Jurassic to Early Cretaceous (184–140 Ma). Pankhurst et al. (2000) recognized episodic eruptions with the first eruptions coinciding with the Karoo and Ferrar LIPs. The province potentially extends into West Antarctica.

Deccan (*DEC*). The Deccan Traps provide one of the most impressive examples of continental flood basalts. It formed by eruptions of primarily tholeiitic basalts over Archean crust over an area of >1.5 Mkm^2 and volume of 8.2 Mkm^3 (Coffin and Eldholm, 1993). The main eruptions straddle magnetic chrons C30*n*, C29*r*, and C29*n* within 1 My around the K–T boundary, as confirmed by Ar/Ar and Re–Os dating (Allegre et al., 1999; Courtillot et al., 2000; Hofmann et al., 2000). The main pulse may have been preceded by a smaller pulse of volcanism at 68–67 Ma (Chenet et al., 2007). An Ir anomaly embedded in the flows suggests that the Chicxulub impact happened while the Deccan Traps were active (Courtillot and Renne, 2003), which is confirmed by stratigraphic work (Keller et al., 2008). Seafloor spreading between India and Seychelles is generally thought to have started in the final stages of flood basalt volcanism (Hooper et al., 2010) at ~63 Ma (Vandamme et al., 1991), although some evidence for an earlier event exists (Collier et al., 2008; Minshull et al., 2008). Unlike older continental flood basalts associated with the breakup of Gondwana, the basalts that are least contaminated by continental lithosphere closely resemble hotspot basalts in oceanic areas with major-element concentrations in agreement with predictions for high-temperature melting (Hawkesworth et al., 1999).

Karoo–Ferrar (*KAR–FER*). The Karoo province in Africa and Ferrar basalts in Antarctica record a volume of 2.5 Mkm^3 combined, which erupted at ~184 Ma (Encarnacion et al., 1996; Minor and Mukasa, 1997), possibly followed by a minor event at 180 Ma (Courtillot and Renne, 2003). The short (<1 My) duration is questioned by Jourdan et al. (2004; 2005), who obtained ages of ~179 Ma for the northern Okavango dike swarm in Botswana and thus inferred long-lived initial activity that propagated northward. More recent zircon U–Pb work on samples collected over an 1100 km range across the Karoo again suggests rapid emplacement (Svensen et al., 2012). In Africa, tholeiitic basalts dominate but some picrites and rhyolites are also exposed (Cox, 1988). The triple-junction pattern of the radiating dike swarm that supplied the Karoo basalt was likely controlled by preexisting lithospheric discontinuities that include the Kaapvaal and Zimbabwe Craton boundaries and the Limpopo mobile belt (Jourdan et al., 2006). The Ferrar province spans an area of ~0.35 Mkm^2 (Elliot and Fleming, 2004) along the Transantarctic Mountains. The two provinces were split by continental rifting and then by seafloor spreading at ~156 Ma.

Madagascar (*MDR*). Widespread voluminous basaltic flows and dikes occurred near the northwestern and southeastern coasts of Madagascar during rifting from India around 88 Ma (Storey et al., 1997). Flood volcanism was probably prolonged as it continued to form the Madagascar Plateau to the south, perhaps 10–20 My later as inferred from the reconstructed positions of Marion hotspot.

North Atlantic Volcanic Province (*NAVP*). The NAVP covers an area of ~1.3 Mkm^2 (Saunders et al., 1997) with an estimated volume of 6.6 Mkm^3 (Coffin and Eldholm, 1993) and is closely linked to continental rifting and oceanic spreading (e.g., Nielsen et al., 2002; White and McKenzie, 1989). Prior to the main pulse of flood volcanism, seafloor spreading was active south of the Charlie–Gibbs Fracture Zone at 94 Ma and propagated northward into the Rockall Trough until the Late Cretaceous near or prior to the earliest eruptions of the NAVP. The early NAVP eruptions occurred as large picritic lavas in western Greenland and Baffin Island (Gill et al., 1992, 1995; Holm et al., 1993; Kent et al., 2004), soon followed by massive tholeiitic eruptions in west and southeast Greenland, the British Isles, and Baffin Island at 61 Ma (2 Mkm^3) as well as in east Greenland and the Faeroes at 56 Ma (>2 Mkm^3) (Courtillot and Renne, 2003; Storey et al., 2007). These initial episodes were succeeded by rifting between Greenland and Europe (recorded by chron 24, 56–52 Ma), continental margin volcanism, and oceanic crust formation, which included the formation of thick seaward-dipping reflector sequences. Spreading slowed in the Labrador Sea at ~50 Ma, stopped altogether at 36 Ma, but continued further to the east on the Aegir Ridge and eventually along the Kolbeinsey Ridge at ~25 Ma, where it has persisted since (Breivik et al., 2006; Mosar et al., 2002). This provides an intriguing suggestion that the presence of hotspots can guide the location of passive ridges during continental breakup.

Many of the volcanic margin sequences erupted subaerially or at shallow depths, suggesting widespread regional uplift during emplacement (Clift and Turner, 1995; Hopper et al., 2003). Uplift in the early Tertiary is documented by extensive erosion and changes in the depositional environments as far as the North Sea basin (e.g., Mackay et al., 2005; Nadin et al., 1997 and references therein). Reconstructions from drill cores elucidate that uplift was rapid and synchronous and preceded the earliest volcanism by >1 My (Clift et al., 1998).

Paraná–Etendeka (*PAR–ET*). Paraná and Etendeka are conjugate volcanic fields split by the breakup of South America and Africa. The Paraná field in South America covers 1.2 Mkm^2 with estimated average thickness of 0.7 km (Peate, 1997). Extensive dike swarms surrounding the provinces suggest the original extent could have been even larger (Trumbull et al., 2004). Tholeiitic lavas and rhyolites cause a bimodal distribution of eruptives and intrusives. Ar/Ar dates indicate that volcanic activity peaked at ~133–130 Ma (Courtillot and Renne, 2003; Renne et al., 1996; Turner et al., 1994), preceded by minor eruptions in the northwest of the Paraná Basin at 135–138 Ma (Stewart et al., 1996). Younger magmatism persisted along the coast (128–120 Ma) and into the Atlantic Ocean, forming the Rio Grande (RIO) and Walvis (WAL) oceanic plateaus.

The Etendeka province covers 0.08 Mkm^2 and is very similar to the Paraná flood basalts in terms of eruptive history, petrology, and geochemistry (Ewart et al., 2004; Renne et al., 1996). Seafloor spreading in the South Atlantic progressed northward, with the oldest magnetic anomalies on the seafloor near Cape Town (137 or 130 Ma). The oldest seafloor near

Paraná is >127 Ma. The formation of onshore and offshore basins suggests a protracted period of rifting well in advance of the formation of oceanic crust and the emplacement of the Paraná basalts (Chang et al., 1992).

7.10.2.3.3 Oceanic LIPs

Caribbean (*CBN*). The Caribbean LIP is a Late Cretaceous plateau, which is now partly accreted onto the continental margin in Colombia and Ecuador. Its present area is 0.6 Mkm^2 with oceanic crust thicknesses ranging from 8 to 20 km. A volume of 4 Mkm^3 of extrusives erupted in discrete events during 91–88 Ma (Courtillot and Renne, 2003). The full range of $^{40}Ar/^{39}Ar$ dates of 69–139 Ma (e.g., Hoernle et al., 2004; Sinton et al., 1997) suggests a protracted volcanic history that is poorly understood. The origin of the volcanism is likely in the eastern Pacific with a subsequent highly mobile evolution. Clogging of eastward subduction in the Caribbean likely led to reversal with westward subduction in the Antilles trench. The accretion of terranes in Ecuador and Colombia, which started in the Late Cretaceous and early Tertiary, led to westward stepping of the subduction zone (Kerr et al., 1997). A 5 km lava succession in Curaçao contains picrites with up to 31 wt% MgO.

Kerguelen (*KER*). The Kerguelen hotspot has a complex history of continental and oceanic flood eruptions, rifting, and prolonged volcanism (**Figure 2**). The breakup of India, Australia, and Antarctica coincided closely with the eruption of the Bunbury basalts in southwest Australia, dated at 123 and 132 Ma (Coffin et al., 2002). The first massive volcanic episode formed the southern Kerguelen Plateau at 119 Ma, the Rajmahal Traps in India at 117–118 Ma, and the lamprophyres in India and Antarctica at 114–115 Ma (Coffin et al., 2002; Kent et al., 2002). The central Kerguelen Plateau and Broken Ridge formed at ~110 Ma and 95 Ma, respectively (Coffin et al., 2002; Duncan, 2002; Frey et al., 2000). These edifices represent the most active period of volcanism with a volcanic area and volume of 2.3 Mkm^2 and 15–24 Mkm^3 (Coffin and Eldholm, 1993), but unlike many other flood basalt provinces, volcanic activity spanned tens of millions of years. Another unusual aspect is the presence of continental blocks within the plateau as suggested from wide-angle seismics (Operto and Charvis, 1996) and geochemical data of xenoliths and basalts from southern Kerguelen Plateau and Broken Ridge (Benard et al., 2010; Frey et al., 2002; Mahoney et al., 1995; Neal et al., 2002). During 82–43 Ma, northward motion of the Indian Plate formed the Ninetyeast Ridge on young oceanic lithosphere, suggesting that the Kerguelen hotspot remained close to the Indian–Antarctic spreading center (Kent et al., 1997; Krishna et al., 2012; Silva et al., 2013; Singh et al., 2011). The influence of the Kerguelen hotspot can be even traced as far as northeastern India (Ghatak and Basu, 2011). At ~40 Ma, the Southeast Indian Ridge formed and separated Kerguelen Plateau and Broken Ridge. Volcanism has since persisted on the northern Kerguelen Plateau until 0.1 Ma (Nicolaysen et al., 2000). Dynamic uplift of the plateau in the Cretaceous is indicated by evidence for a subaerial environment, but subsidence since then appears to be similar than for normal oceanic crust (Coffin, 1992).

Ontong Java Plateau (*OJP*). Overviews of the origin and evolution of the Ontong Java Plateau are provided by Fitton et al. (2004) and Neal et al. (1997). This plateau extends across 1.5–2 Mkm^2, has crust as thick as 36 km, and has volume estimated at ~50 Mkm^3 (Gladczenko et al., 1997; Korenaga, 2011). The majority of the basalts were erupted in relatively short time (~1–2 My) near 122 Ma as revealed by Ar/Ar (Tejada et al., 1996) and Re–Os isotope (Parkinson et al., 2002), paleomagnetic (Tarduno et al., 1991), and biostratigraphic data (Sikora and Bergen, 2004). Basalts recovered from the Ocean Drilling Program (ODP) Site 803 and the Santa Isabel and Ramos Islands indicate a smaller episode of volcanism at 90 Ma (Neal et al., 1997). Age contrasts between the surrounding seafloor and OJP suggest it erupted near an MOR. Furthermore, rifting is evident on some of its boundaries. Accordingly, Taylor (2006) and Chandler et al. (2012) suggested that the OJP is only a fragment of what was originally an even larger LIP that included Manihiki (MAN) and Hikurangi (HIK) Plateaus. This idea is supported by geochemical similarities between OJP and Manihiki lavas (Hoernle et al., 2010). Given recent evidence for partial subduction of HIK beneath New Zealand, current estimates for the extent of the combined LIP should be regarded as lower bounds (Reyners et al., 2011). The estimated original size of the combined LIP makes it, by far, the largest of all flood basalt provinces in the geologic record, approaching an order of magnitude larger in volume than any continental flood basalt.

The plateau has been sampled on the tectonically uplifted portions in the Solomon Islands (e.g., Petterson, 2004; Tejada et al., 1996, 2002) and by ocean drilling (Mahoney et al., 2001). The volcanics are dominated by massive flows of tholeiitic basalts with low potassium contents. Petrologic modeling is consistent with primary magmas formed by 30% melting of a peridotitic source (Chazey and Neal, 2004; Fitton and Godard, 2004; Herzberg, 2004b), which would require a hot (1500 °C) mantle source beneath thin lithosphere. The low volatile content of volcanic glasses (Roberge et al., 2004, 2005) and Pb, Sr, Hf, and Nd isotope characteristics (Tejada et al., 2004) resemble MORB.

Besides its gigantic volume, another enigmatic aspect is that the OJP appears to have erupted below sea level with little evidence for uplift (Coffin, 1992; Ingle and Coffin, 2004; Ito and Clift, 1998; Korenaga, 2005; Roberge et al., 2005). This observation has to be explained by any successful model for the origin of Ontong Java.

Shatsky Rise–Hess Rise (*SHA–HES*). Shatsky Rise is one of the large Pacific oceanic plateaus with an area of 0.48 Mkm^2, a thickness of 9–30 km, and a volume of 4.3 Mkm^3 (Korenaga and Sager, 2012; Sager et al., 1999). The most voluminous portion of the plateau extends from its center to the SW (**Figure 1**). This whole portion has recently been proposed to be made up of a single volcanic edifice with very low slopes, something that would probably make it the largest volcano on Earth (Sager et al., 2013). Initial eruption at Shatsky is associated with a jump of the triple junction between the Pacific, Izanagi, and Farallon Plates toward the plateau (Nakanishi et al., 1999; Sager, 2005) near 145 Ma (Mahoney et al., 2005). Subsequently, volcanism progressed northeast together with the triple junction (which migrated with repeated jumps as indicated by seafloor magnetic lineation) until ~128 Ma. Thus, Shatsky appears to show short-lived age-progressive volcanism on timescales much like many smaller volcano groups (e.g., Society).

Volcanism, however, may have continued with a renewed pulse starting some 10–20 My later with the formation of the Hess Rise, which is comparable in area to Shatsky. Age constraints on Hess Rise are poor due to the lack of sampling and its location on Cretaceous quiet zone seafloor. The possible coincidence of both plateaus at an MOR suggests a dynamic linkage between their formation and seafloor spreading. Another notable aspect is that Pb and Nd isotope compositions for Shatsky are indistinguishable from those of the present-day East Pacific Rise (Mahoney et al., 2005; Sager, 2005).

7.10.2.3.4 Connections to hotspots

The possible links between hotspots and LIPs are important for testing the mechanism of origin for both phenomena, with particular regard to the concept of a starting mantle plume head and trailing, narrower plume stem. While linkages are clear for some cases, a number of proposed connections are obscured by ridge migrations or breakup of the original LIP. In the succeeding text, we list the connections of hotspots to LIPs, in approximate order of decreasing clarity.

At least six examples have strong geographic, geochronological, and geochemical connections between hotspot volcanism and flood basalt provinces. These are (1) Iceland and the North Atlantic Volcanic Province, including Greenland, Baffin Island, Great Britain volcanics, and Greenland–United Kingdom (Faeroe) Ridge (Saunders et al., 1997; Smallwood and White, 2002); (2) Kerguelen, Bunbury, Naturaliste, and Rajmahal (E India) and Broken Ridge and Ninetyeast Ridge (Kent et al., 1997); (3) Réunion and Deccan (Roy, 2003), W. Indian, Chagos–Laccadive, Mascarene, and Mauritius; (4) Marion and Madagascar (Meert and Tamrat, 2006; Storey et al., 1997); (5) Tristan da Cunha and Paraná, Etendeka, Rio Grande, and Walvis Ridge (Peate, 1997); and (6) the Galápagos and Caribbean (Feigenson et al., 2004; Hoernle et al., 2004).

In addition to these six examples, a tentative link exists between Yellowstone and the early eruptive sequence of the Columbia River basalts (CRBs) based on geochemistry (Dodson et al., 1997). The paleogeographic connection is somewhat indirect since main CRB eruptions occurred up to 500 km north of the hotspot track. This offset may suggest a mechanism for the formation of the flood basalts that is independent of the Yellowstone hotspot (Camp and Hanan, 2008; Hales et al., 2005; Long et al., 2012). Yet, the earliest manifestation of the CRB (at Steens Mountain) is indeed close to the proposed paleolocation of the Yellowstone hotspot (Hooper, 1997; Hooper et al., 2002), implying a south-to-northward propagation of CRB volcanism (Coble and Mahood, 2012) and supporting the suggestion that melts may have been forced sideways from their mantle source by lithospheric structure (Thompson and Gibson, 1991). In turn, it has been suggested that the lithospheric structure forced the Yellowstone plume away from the flood basalt province after ~12 Ma (Shervais and Hanan, 2008).

For other flood basalt provinces, the links are less clear – in part due to the lack of data. The Bouvet hotspot has been linked to the Karoo–Ferrar; the Balleny hotspot may be connected to the Tasmanian province or the Lord Howe Rise (Lanyon et al., 1993); and the Fernando hotspot has been linked to the Central Atlantic Magmatic Province. Only three Pacific hotspots possibly link back to LIPs: Louisville–OJP, Easter–Mid-Pac, and Marquesas–Hess–Shatsky (Clouard and Bonneville, 2001). The Louisville–OJP connection is highly tenuous: Kinematic arguments against such a link have been made (Antretter et al., 2004), and geochemical differences between the oldest Louisville seamounts and OJP would require distinct chemistry between plume head and tail or a difference in melting conditions (Mahoney et al., 1993; Neal et al., 1997). Nevertheless, some hotspots such as Hawaii, Bowie–Kodiak, and Cobb terminate at subduction zones, so any record of a possible connection to an LIP has been destroyed.

It is interesting to note that the strongest connections involve flood basalt provinces near continental margins. The possible exceptions are Kerguelen and Marion–Madagascar, which also involve continental lithosphere, as well as the tentative links in the Pacific. The presence of continents may indeed play a large role in the origin of flood basalt volcanism (Anderson, 1994b).

7.10.2.3.5 Connections to climate crises and mass extinction

The eruption of LIPs has been a leading explanation for a number of climate crises and mass extinctions. This volcanic cause of mass extinctions poses an alternative to extraterrestrial causes such as impacts by asteroids or comets, which are probably best documented to be associated with the K–T mass extinction (Schulte et al., 2010). The connection between mass extinctions and LIPs is generally justified by the coincidence of LIP formation and extinctions as well as the ability of rapid and massive extrusions to dramatically change climate and the chemistry of the atmosphere and oceans. The episodicity of mass extinction events and their correlation with LIPs lead to the speculation that the slow convection and heat release from the solid Earth control evolutionary patterns at the Earth's surface (e.g., Courtillot and Olson, 2007). Increasingly, a connection is seen between flood basalts and other transient episodes in the global climate system. Aside from direct effects on ocean and atmospheric chemistry due to volcanic outgassing, the igneous intrusions associated with LIPs can lead to significant release of CO_2 and methane (i.e., major greenhouse gases) from metamorphic interactions with coal, evaporates, and other sedimentary rocks (e.g., Ganino and Arndt, 2009; Retallack, 2013). Such links are also suggested for flood basalts that occur during continental breakup, as illustrated by the correlation between North Atlantic flood basalts with the Paleocene–Eocene thermal maximum at ~55 Ma (Storey et al., 2007) and the Caribbean and Madagascar flood basalts with the Late Cretaceous anoxic event (Kuroda et al., 2007). Interestingly, while the short-term effect of CO_2 release from flood basalt volcanism is likely to lead to global warming, the longer-term effect may be global cooling due to more efficient silicate weathering that acts as a carbon sink (Schaller et al., 2011, 2012).

Significant recent work has refined early indications that mass extinctions coincide with LIP formation. Indications have indeed been strengthened for the four major extinctions that pair up with LIP formation since the Permian: the end of Cretaceous extinction with the Deccan Traps at ~65 Ma, the end of Triassic extinction with the CAMP at ~200 Ma, the end of Permian extinction with the Siberian Traps at ~250 Ma, and the end of Guadalupian (an epoch in the Permian) extinction with the Emeishan at ~258 Ma (for reviews, see Courtillot and

Renne, 2003; Bryan et al., 2010). In some cases, the suggested temporal coincidence has been questioned because of inherent uncertainties in magnetostratigraphy and geochronological techniques (Whiteside et al., 2007). Yet, future geochronological work using high-precision dating tools is expected to resolve these discrepancies. For example, Blackburn et al. (2013) demonstrated using a new zircon U–Pb methodology that the end of Triassic extinction coincides with four episodes of magma release in the CAMP and atmospheric changes that have occurred over a short period of only 0.6 My. A combination of multiple geochronological techniques further confirms the connection between CAMP and the end of Triassic extinction (Deenen et al., 2010).

The LIP–extinction connection is further strengthened by a better quantification and qualification of the duration and chemical output of eruptions, as well as a better understanding of the impact of flood volcanism on the global climate system (e.g., Black et al., 2012; Kidder and Worsley, 2010; Payne and Clapham, 2012; Self et al., 2006) and ocean geochemistry (Brand et al., 2012). This understanding is enhanced by a better physical description of the interaction between magmas and sediments (Aarnes et al., 2010; Ogden and Sleep, 2012). Some of this work indicates intriguing differences in magmatic composition between LIPs, such as the anomalously high concentration in sulfur and halogens observed in the Siberian Traps flood basalts (Black et al., 2012), which may have significantly contributed to the dramatic deterioration of the environment at the end of the Permian. The anomalous impact of the Siberian Traps is generally attributed to the intrusion of the basalts into a deep sedimentary basin with significantly higher gas release (Ganino and Arndt, 2009; Svensen et al., 2009) although differences in the source have also been noted (Carlson et al., 2006; Sobolev et al., 2009, 2011).

7.10.2.4 Geochemistry and Petrology of OIB

7.10.2.4.1 Isotope geochemistry

Studying the geochemistry of MORB and basalts from oceanic hotspots and melting anomalies (often referred to as ocean island basalts, OIBs) has played a central role in understanding mantle dynamics and heterogeneity. Isotope ratios of key trace elements reflect long-term (10^2–10^3 My) concentration between parent and daughter elements. As they remain unaffected by mantle metasomatism and melting, they have been used to 'fingerprint' source materials. We will focus on three commonly used ratios: $^{87}Sr/^{86}Sr$, $^{206}Pb/^{204}Pb$, and $^{3}He/^{4}He$. High (or low) $^{87}Sr/^{86}Sr$ is associated with long-term enrichment (or depletion) in highly incompatible elements relative to moderately incompatible elements (e.g., Rb/Sr). High (or low) $^{206}Pb/^{204}Pb$ ratios are associated with mantle with a long-term high (or low) U/Pb. The $^{3}He/^{4}He$ ratio is a measure of the amount of primordial ^{3}He, which has been lost to space during the Earth's evolution via degassing but not generated in the mantle, in contrast to radiogenic ^{4}He.

MORB is characterized by low values of, and small variability in, the ratios mentioned earlier. **Figure 11** shows frequency distributions of MORB and OIB from the three major oceans. The data are from http://www.petdb.org/; all data are included so there is no filtering due to affiliation with hotspots or any tectonic features. One standard deviation about the median for all three datasets combined defines the MORB range for $^{87}Sr/^{86}Sr$ of 0.7025–0.7033, for $^{206}Pb/^{204}Pb$ of 17.98–18.89, and for $^{3}He/^{4}He$ of 7.08–10.21. Excluding hotspot-influenced MORB would further reduce these ranges (e.g., to a narrow range in $^{3}He/^{4}He$ of 7.68–9.82 (Graham, 2002)). In contrast, OIBs have a larger variability, extending from MORB values to much higher values (http://georoc.mpch-mainz.gwdg.de/georoc/; **Figure 11**). By comparing the median hotspot compositions (solid bar) with the MORB ranges, it becomes clear that, with few exceptions, the hotspots and melting anomalies have compositions that are distinguishable from MORB by at least one of the three isotope ratios (see also **Table 1**). This result appears to be independent of the duration of age progression (e.g., Tristan vs. Marquesas), the presence or absence of a swell (Hawaii vs. the Canaries), or even volcanic flux (Kerguelen vs. Pukapuka and Foundation).

There are at least four possible exceptions: Shatsky Rise (not shown in **Figure 11**), Cobb, Bowie–Kodiak, and the Carolines. Helium isotopes are not yet available at these locations, but $^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$ compositions for each case fall within or very near to the MORB range. These examples span a wide range of forms, from a small, short-lived seamount chain (the Carolines), to longer-lived, age-progressive volcanism (Bowie–Kodiak and Cobb), to an oceanic LIP (Shatsky). The possibility that these cases are geochemically indistinguishable from MORB may have far-reaching implications for mantle processes and composition and thus clearly motivates further sampling and analysis.

7.10.2.4.2 Major-element geochemistry

In contrast to isotopic ratios, major-element signatures of OIB lavas do not directly reflect that of the mantle source. Major-element concentrations in primary magmas (i.e., magmas that segregated from the mantle melting zone without further modification by shallow processes) are sensitive to the depths and extents of melting. Moreover, they are affected by magma mixing, interaction with the wall rock, and fractional crystallization during ascent. The effects of fractional crystallization are not only large but also relatively well understood and thus can be corrected for when estimating the composition of primary magmas. Alternatively, primary magma compositions can be inferred from melt inclusions or olivine phenocrysts that are estimated to have been in equilibrium with the primary magma when they were formed. However, such analyses are laborious and costly compared to whole rock analysis and hence form just a small portion of the global dataset.

Despite the difficulties mentioned earlier, major-element signatures of OIB and MORB can be clearly distinguished in properly selected and corrected datasets of whole rock analyses. For example, these datasets are usually restricted to samples with MgO contents between >5 wt% (or even >10 wt%) and $< \sim 15$ wt% in order to exclude lavas that have experienced significant fractional crystallization of clinopyroxene (Dasgupta et al., 2010; Jackson and Dasgupta, 2008; Pilet et al., 2008). Fractional crystallization of olivine instead is routinely corrected for by retroactive incremental addition or subtraction of olivines that have been in equilibrium with the corresponding liquid. Similar to isotope datasets, such corrected major-element datasets show that OIB global trends overlap with MORB, but display a much greater variability in all oxides and oxide ratios (**Figure 12**). In contrast to MORB,

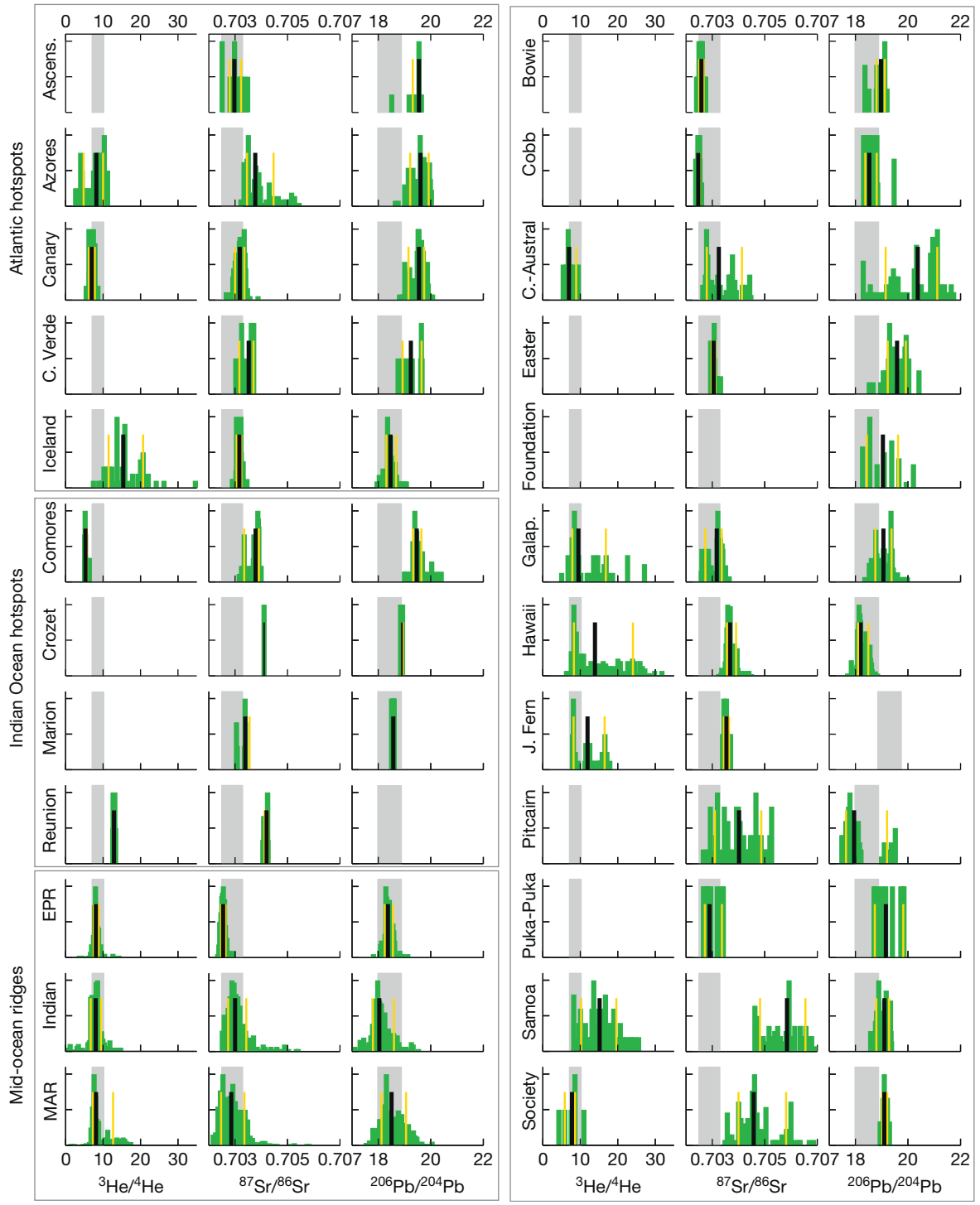


Figure 11 Frequency distributions (green, normalized by maximum frequency for each case) of isotope measurements taken for each of the oceanic hotspots shown and for MORs (lower left). Bold black lines mark median values, and yellow lines encompass 68% (i.e., one standard deviation of a normal distribution) of the samples. Light gray bars denote the range of values encompassing 68% of all of the MORB measurements (sum of the three ridges shown). Most of these data are from the GEOROC database (http://georoc.mpch-mainz.gwdg.de/georoc/) with key references for $^{3}He/^{4}He$ data given in Ito and Mahoney (2006). Data for Pukapuka are from Janney et al. (2000) and for the Foundation chain from Maia et al. (2000).

OIB commonly follows an alkalic differentiation trend (except for shield lavas from Hawaii, the Galápagos, and Iceland) due to systematically higher $K_2O + Na_2O$ at a given SiO_2 in the primary magmas. This subset of alkalic OIB is hence nepheline-normative, that is, highly undersaturated in SiO_2. They also display higher concentrations of K_2O, MgO, FeO_X, and TiO_2, as well as lower concentrations of Al_2O_3 at a given SiO_2 (e.g., Pilet et al., 2008), than tholeiitic OIB or even MORB. These characteristics are important constraints for the makeup of the OIB mantle source (see **Section 7.10.3.6**).

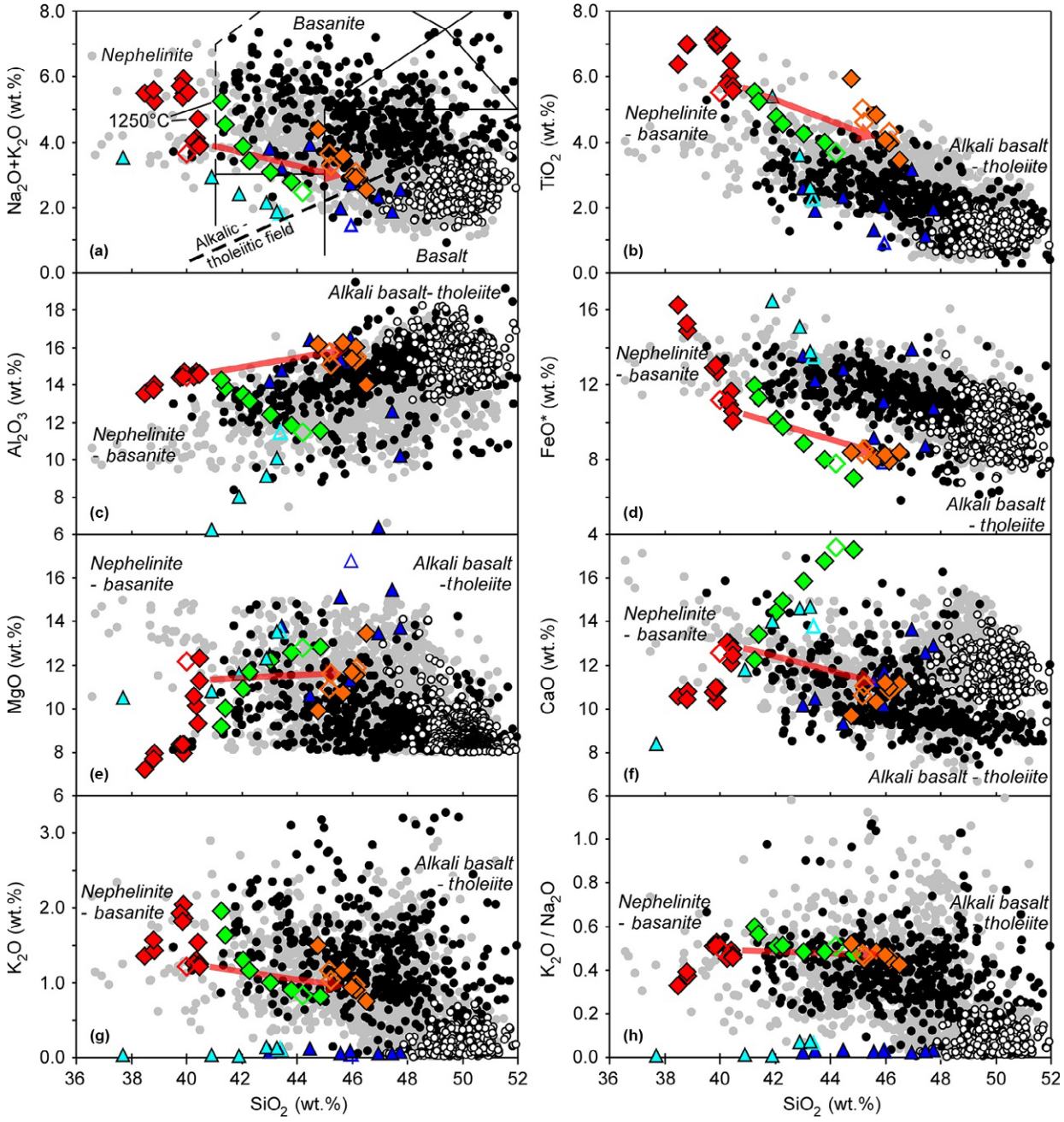


Figure 12 From Pilet et al. (2008) with minor modifications. Select major-element concentrations (a–g) and oxide ratios (h) of basalt samples and experimental melts. Solid gray circles, OIB; solid black circles, continental intraplate basalts; open circles, MORB. Solid diamonds are experimental hornblenditic liquids from three different starting compositions (open diamonds) at various experimental conditions (Pilet et al., 2008); triangles are experimental liquids derived from silica-deficient pyroxenites with (dark blue; Dasgupta and Hirschmann, 2006) or without (light blue; Hirschmann et al., 2003; Kogiso et al., 2003) the addition of CO_2.

7.10.2.5 Mantle Seismic Anomalies

7.10.2.5.1 Global seismic studies

Seismic wave propagation in the mantle is generally slowed by elevated temperature, volatile content, and the presence of melt (e.g., Anderson, 1989). In addition, it is affected by the content of mafic materials, such as eclogite (Xu et al., 2008; Zhang and Green, 2007). Seismology is therefore the primary geophysical tool for probing the physical signature of hotspots and melting anomalies and for identifying their depths of origin.

Seismic tomography uses seismic delay times and waveform deviations to map velocity anomalies relative to a particular background model. Applications of these techniques have seen wide popularity as velocity variations can be directly related to chemical and thermal variations in the Earth. Yet, it should be considered that any tomographic image is in fact a

model of the Earth that is strongly dependent on the method of inversion, choice of damping parameters, and background model and can be quite sensitive to the lack of global coverage of earthquakes and seismometers, as well as standard consequences of wave propagation such as wave-front healing or those associated with finite-frequency effects.

Global seismic tomography has been able to confidently image the remnants of slabs as high-velocity and presumably cold anomalies (e.g., Bijwaard et al., 1998; Grand et al., 1997). The hot mantle upwellings hypothesized to be associated with hotspots are instead expected to be narrow, localized, and much less voluminous than slabs. Any associated low-velocity anomalies may further be obscured by the effects of wave-front healing. Combined, these effects should make plumes very difficult to be imaged with seismic tomography, especially in the lower mantle (Hwang et al., 2011; Nolet et al., 2007; Ritsema and Allen, 2003; Styles et al., 2011). Structures in the lower mantle that are most robustly imaged and likely associated with upwelling are the (1000s of km across) LLSVPs below the South Pacific and African Superswell regions (e.g., Breger et al., 2001; Ni et al., 2005; Trampert et al., 2004; van der Hilst and Karason, 1999). The sharp edges of these anomalies (Ni et al., 2002; To et al., 2005) and their reproduction in dynamic models (McNamara and Zhong, 2005; Tan and Gurnis, 2007) suggest that these features are anomalous in terms of both composition and temperature (see also **Chapter 7.11**).

The use of global tomography to probe the existence and nature of smaller-scale (100 s of km across) upwellings associated with individual hotspots continues to be a challenge, but substantial progress has been made with improved resolution in global seismic models (e.g., Boschi et al., 2007; Montelli et al., 2006; Nataf, 2000; Nolet et al., 2007; Ritsema and Allen, 2003; Zhao, 2007 and references therein). Numerous hotspots have been shown to overlie low-velocity anomalies in the upper mantle, and a subset of these anomalies have been argued to extend into the deep lower mantle based on visual inspection (Montelli et al., 2006; Zhao, 2007; **Figure 13**) or on various statistical tests (Boschi et al., 2007, 2008). Two research groups (Boschi et al., 2007; Zhao, 2007) emphasize that many lower mantle, plume-like anomalies are best explained by plumes, which tilt and whose sources migrate due to flow of the ambient mantle. In addition, evidence shows that lower mantle plumes tend to originate from around the edges of the LLSVPs beneath the South Pacific and Africa (Boschi et al., 2007; Thorne et al., 2004; **Figure 13(e)**).

The passage of mantle plumes from the lower mantle into the upper mantle can also be detected by measuring the depths of the seismic discontinuities bounding the mantle transition zone. Excess temperature is expected to cause the phase change near 410 km with a positive Clapeyron slope to occur deeper and the phase change near 660 km with a negative Clapeyron slope to occur shallower. In both the cases, the olivine system is assumed to dominate the phase changes (Helffrich, 2000; Ito and Takahashi, 1989). Hot mantle plumes that pass through both phase changes should therefore be associated with unusually thin transition zones. Some global-scale studies suggest this is indeed the case (e.g., Lawrence and Shearer, 2008; Li et al., 2003; Schmerr et al., 2010; Tauzin et al., 2008). In some cases, most of the variations in transition zone thickness to topography are suggested at 660 km (Lawrence and Shearer, 2008; Schmerr et al., 2010), whereas many others attributed most of the variations to topography at 410 km (Li et al., 2003; Tauzin et al., 2008). The evidence for the transition near 660 km tending to remain flat or deepen – not shoal – in some regions where the mantle is expected to be unusually hot could indicate that the phase change effects are locally dominated by the garnet system (postgarnet to perovskite, with a positive Clapeyron slope) (Cao et al., 2011; Houser and Williams, 2010; Tauzin et al., 2008).

Results about the depth of origin of plumes beneath many of the Earth's major hotspots are beginning to show some consistencies. Between the studies of Montelli et al. (2006) and Zhao (2007), 19 hotspots are associated with plumes originating in the deep lower mantle, but only four (Tahiti, Kerguelen, Hawaii, and Cape Verde) are recognized as such in both of the studies (**Figure 14**). These four hotspots also show continuity of low-velocity material extending through most of the mantle according to Boschi et al.'s (2007) objective analysis (i.e., normalized vertical extent >0.5 in **Figure 14**). An additional seven other hotspots (East Africa/Afar, Réunion, Macdonald/Cook, Samoa, the Canaries, the Azores, Iceland, and Easter) identified as deep plumes by either Montelli et al. (2006) or Zhao (2007), but not by both, are also likely to extend through most of the mantle according to Boschi et al. (2007). Of the 19 identified by Montelli et al. (2006) and Zhao (2007), 6 hotspots (Tahiti, Samoa, the Canaries, the Azores, Iceland, and Hawaii) are further associated with anomalously thin mantle transition zones, but 1(Easter) is associated with a transition zone that is as thick or thicker (246 ± 3 km) than the global average (242 ± 2 km) (Courtier et al., 2007). Six of these nineteen hotspots (Easter, Hawaii, Iceland, Réunion, East Africa, and Louisville) were identified by Courtillot et al. (2003) as being 'primary plumes' from the deep mantle on the basis of their association with a buoyancy flux $>10^3$ kg s^{-1}, a flood basalt province, and magmas with high ^{3}He/^{4}He (see also purple squares in **Figure 7**).

Along these lines, at least three research groups agree that ~11 or so hotspots originate as hot mantle upwellings rising from the deep lower mantle (**Figure 14**): Tahiti, East Africa, Réunion, Samoa, Kerguelen, the Canaries, the Azores, Iceland, Hawaii, Cape Verde, and Easter Island. A critical shortcoming is that the resolution of seismic tomography in the lower mantle is probably insufficient to seismically resolve slow mantle upwellings with lateral dimensions of hundreds of kilometers (Styles et al., 2011). The regions of the lower mantle most reliably imaged as being seismically slow have widths of 1000 km or more (Boschi et al., 2007). This leads to the unresolved question whether any lower mantle plumes (100 s of km wide) originate individually from the core–mantle boundary (CMB) or whether all lower mantle plumes rise out of much broader (1000s of kilometer wide) dome-like LLSVPs (Courtillot et al., 2003; Schubert et al., 2004; Torsvik et al., 2006). Of course, the lack of visibility of a narrow plume in the lower mantle may also just indicate its absence and an origin within or near the transition zone (Courtillot et al., 2003; Cserepes and Yuen, 2000).

7.10.2.5.2 Seismic studies of major hotspots

Regional studies of the seismic structure of the mantle beneath hotspots are often at higher resolution than global studies. We will discuss the main results of these in geographic order.

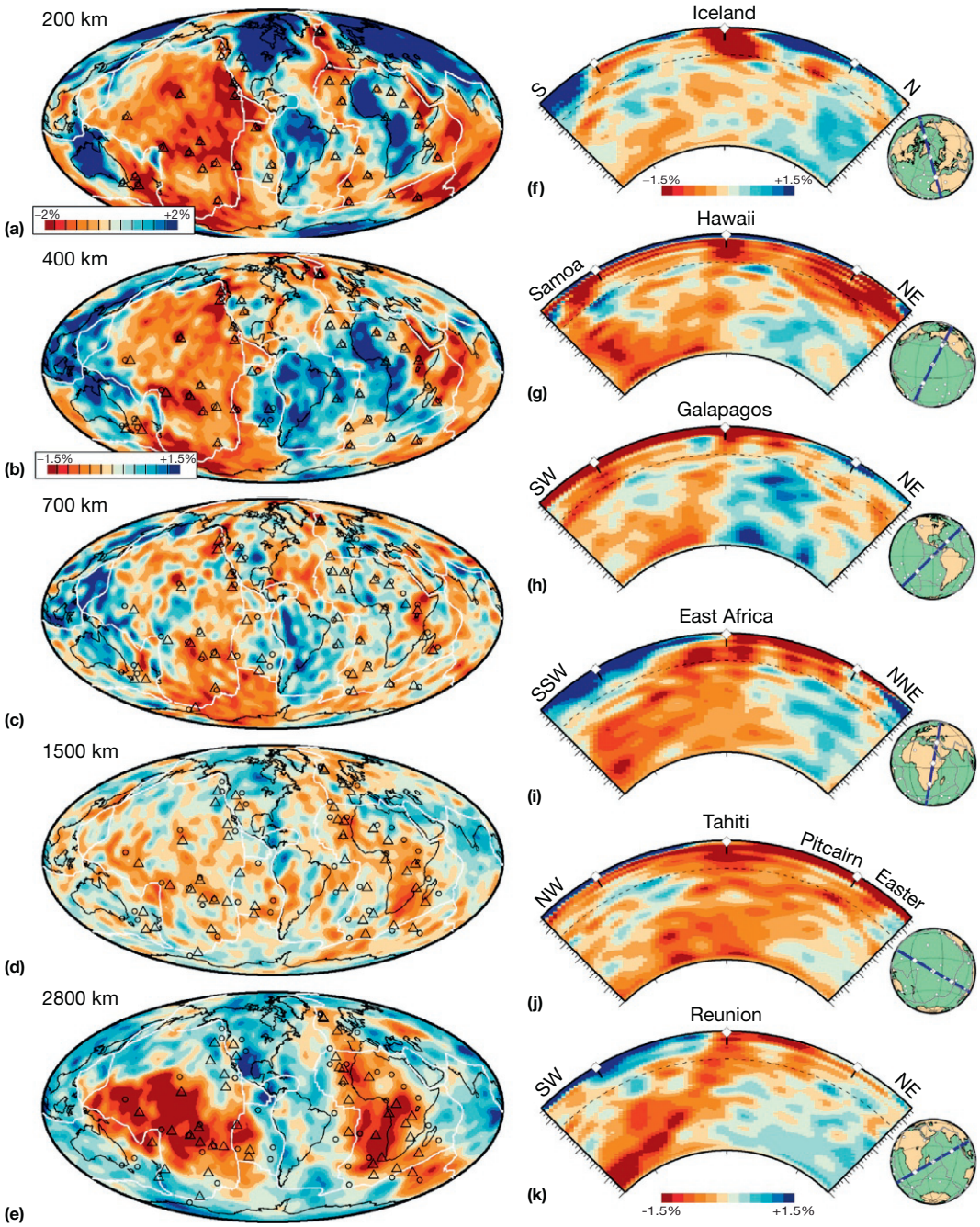


Figure 13 (a)–(e) Maps of shear-wave velocity anomalies at different depths (as denoted) in tomography model S40RTS (Ritsema et al., 2011) (all have the same color scale except (a)). Circles mark locations of hotspots (Steinberger, 2000). Estimated plume locations at the different depths (triangles) vary because plumes are modeled as swaying with predicted mantle flow and rising from source locations at the CMB, which also move with the mantle (Boschi et al., 2007; Steinberger and Antretter, 2006). (f)–(k) 90° cross sections beneath several hotspots with low-velocity zones in the upper mantle that appear to extend into the lower mantle. Insets show location and orientations of the cross sections. Diamonds mark three points along each cross section for orientation.

Iceland: Regional seismic studies have confidently imaged a body of anomalously slow seismic wave speeds in the upper mantle beneath Iceland. Conventional ray theory was used to first image the anomaly (Allen et al., 1999, 2002; Foulger et al., 2001; Wolfe et al., 1997). A subsequent study using an improved finite-frequency technique (Allen and Tromp, 2005) resolves the feature to be roughly columnar with lateral dimensions of 250–300 km with peak P- and S-wave anomalies of −2.1% and −4.2%, respectively (Hung et al., 2004; **Figure 15**). Recent studies using full-waveform tomography (Rickers et al., 2013) or Rayleigh waves and local earthquakes (Li and Detrick, 2006; Yang and Shen, 2005) confirm these high amplitudes, which most likely require a combination of excess temperature and melt.

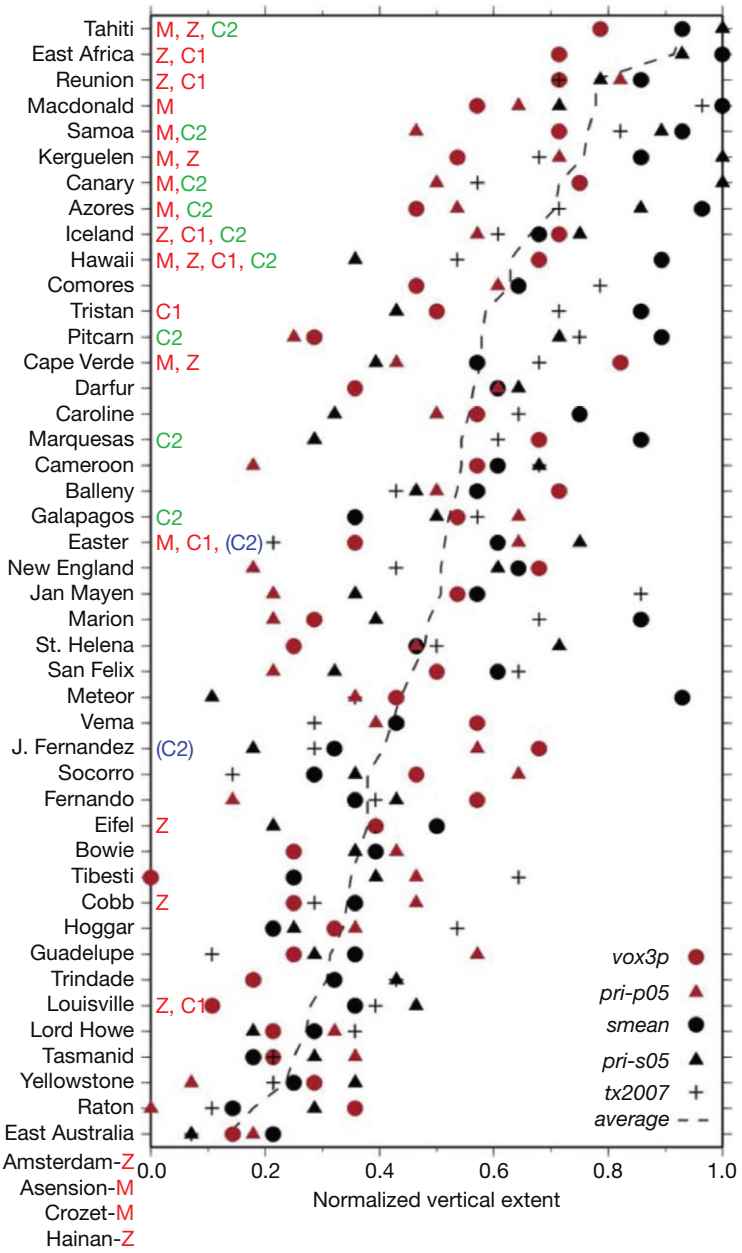


Figure 14 Continuity of plumes in seismic tomography models; modified from Boschi L, Becker TW, and Steinberger B (2007) Mantle plumes: Dynamic models and seismic images. *Geochemistry, Geophysics, Geosystems* 8. Each point in a seismic tomography model was examined along the predicted trajectory of a plume that rises from the CMB, is deflected by mantle flow, and surfaces at a given hotspot (listed above). If the point was seismically slow by $\geq|1.5\%|$, it was assigned a value 1.0; otherwise, it was assigned a 0.0. The points were then averaged along the plume trajectory to yield 'normalized vertical extents.' A value of 1.0 indicates that the plume-like anomaly in the tomography model extends across the whole mantle. A value near 0.0 indicates the plume-like anomaly is restricted to shallow depths. Symbols denote values from three S-wave models (tx2007 (Simmons et al., 2006), smean (Becker and Boschi, 2002), and pri-s05 (Montelli et al., 2006) and two P-wave models (pri-p05 (Montelli et al., 2006) and vox3p (Boschi et al., 2007)). Dashed profile shows the average of all the points. Beyond this analysis, red letters indicate studies ('C1,' Courtilot et al. (2003); 'M,' Montelli et al. (2006); and 'Z,' Zhao (2007)), which identify a deep lower mantle origin for a given hotspot. 'C2' (green) and '(C2)' (blue) indicate hotspots overlying a mantle transition zone that is anomalously thin or normal, respectively (Courtier et al., 2007).

In addition, studies of surface and S-waves reveal seismic anisotropy in the Icelandic upper mantle. Shear-wave splitting of the SKS phase measures the integrated effects of seismic anisotropy along nearly vertical ray paths across the mantle and is detected at a number of locations on Iceland with the fastest S-waves generally being polarized NW–SE (Bjarnason et al., 2002; Li and Detrick, 2003; Xue and Allen, 2005). Azimuthal anisotropy of Rayleigh waves has also been measured. Along the zones of active rifting and volcanism on Iceland, this anisotropy is strong at shallow depths <50 km with fast directions NNE–SSW and weak when somewhat deeper (Li and Detrick, 2006). West of the rift zones, fast

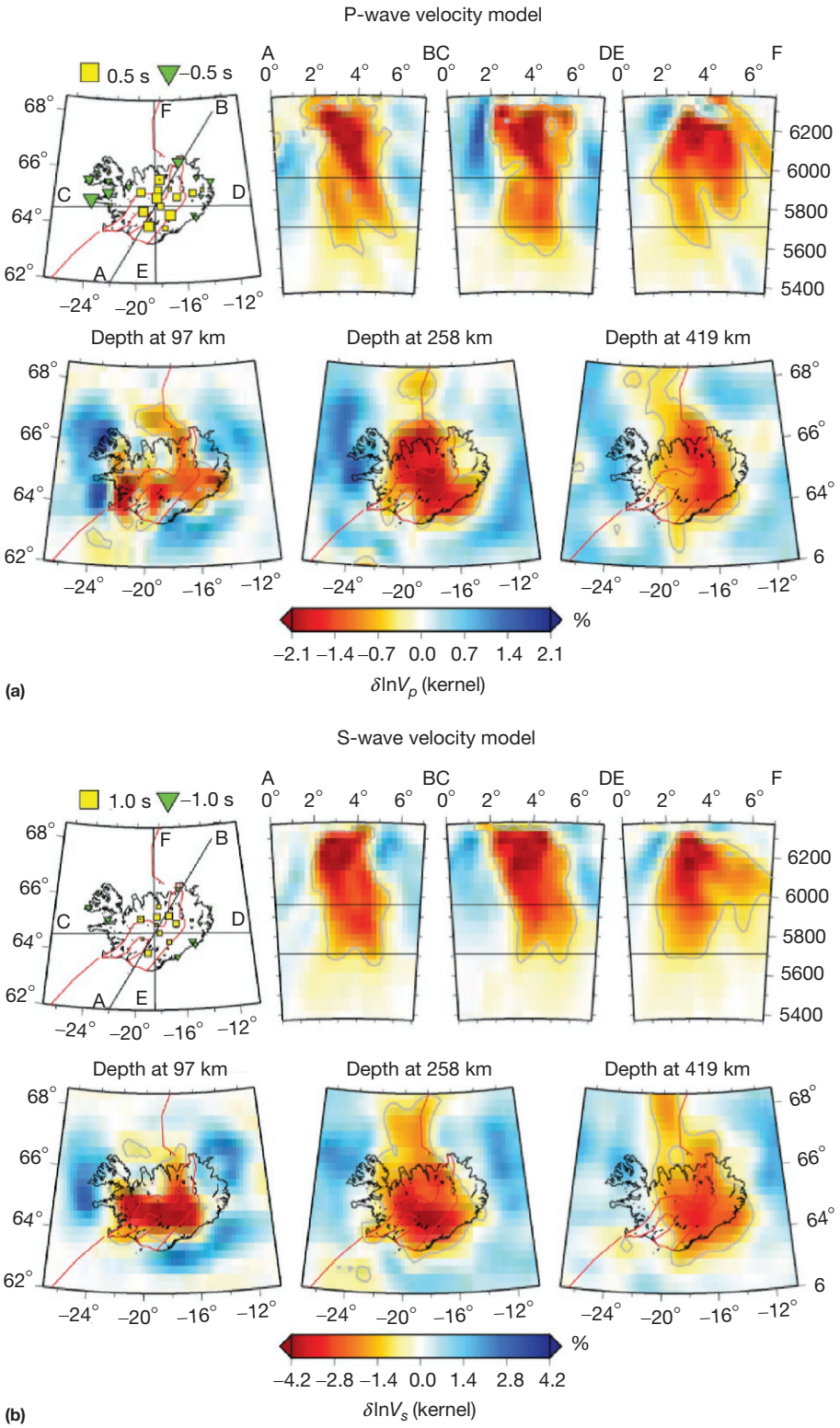


Figure 15 Tomographic inversions of the mantle below the Iceland hotspot (Hung et al., 2004) for (a) P-waves and (b) S-waves. Top row shows vertical cross sections along three transects (see insets). Bottom row shows horizontal sections at depths as labeled.

directions are instead oriented NW–SE at shallow depths and more variably at depth. Taken together, the S-wave splitting and surface-wave results are not straightforward to interpret and lead to different conclusions about the cause of seismic anisotropy beneath Iceland. Li and Detrick (2006) interpreted the NNE–SSW Rayleigh wave anisotropy as being caused by a preferential alignment of melt channels or by lattice-preferred orientation (LPO) of olivine induced by mantle flow that roughly parallels the NNE–SSW-trending Northern Volcanic Rift Zone and South Iceland Seismic Zone. In contrast, Xue and Allen (2005) interpreted the S-wave splitting to indicate LPO due to flow from the center of the hotspot in eastern Iceland to the Kolbeinsey Ridge north of Iceland.

The vertical extent of the Iceland mantle plume has been examined in a variety of studies. Receiver function studies show that the Iceland plume extends well below 410 km as shown by the thinning of the transition zone beneath Iceland (Du et al., 2006; Shen et al., 1998, 2002). While Shen et al. (1998, 2002) showed evidence for both a deepening of the 410 and a shoaling of the 660 km discontinuity SE of Iceland, Du et al. (2006) argued that the 660 km is instead flat. The precise nature of both discontinuities is important in determining whether the Iceland seismic anomaly initiates in the lower mantle. A recent regional full-waveform tomography model (Rickers et al., 2013) is in agreement with Shen et al. (1998; 2002) and suggests that the anomaly is tilted to the SE and extends into the lower mantle. At least three global tomography models are consistent with such a deep-rooted anomaly (Bijwaard and Spakman, 1999; Boschi et al., 2007; Zhao, 2007; see also **Figure 14**). A separate study identifies an ultralow-velocity zone near the CMB below Iceland (Helmberger et al., 1998). An updated global tomography model S40RTS (Ritsema et al., 2011) shows that the lower mantle is seismically slow in a broad zone beneath Iceland but with weak and variable amplitude, suggesting difficulty in resolving the continuity of the Iceland plume from the CMB (**Figure 13(f)**). A more robust determination of the depth of origin of the Iceland plume probably requires a regional seismic experiment spanning a wider geographic area than just the island of Iceland.

The Azores: Evidence for anomalously hot mantle beneath the Azores hotspot comes from a locally broad area of anomalously slow surface-wave speeds identified in the upper 200 km of the mantle, which appears as a perturbation of the generally low-velocity structure extending along the Mid-Atlantic Ridge (Pilidou et al., 2004). Finite-frequency body-wave tomography reveals an irregularly shaped anomaly of low P-wave speeds beneath the Azores archipelago in the shallowest 200 km, which slants northeast and downward to ~400 km depth (Yang et al., 2006).

Cape Verde: Two groups find conflicting results for the mantle structure beneath the Cape Verde hotspot, while both use receiver function methods. The first group finds that the uppermost mantle beneath Cape Verde has anomalously high seismic velocity and low density (Lodge and Helffrich, 2006) and that the transition zone has a normal thickness (Helffrich et al., 2010). The former result was interpreted as revealing a layer of depleted residue from hotspot melting and the latter as indicating that the mantle upwelling sustaining the hotspot originates in the upper mantle. In contrast, Vinnik et al. (2012) found low seismic velocities and normal densities in the shallow upper mantle, as well as a ~30 km thinning of the transition zone, due to a deepening of the 410 km discontinuity and a shoaling of the 660 km discontinuity. These results are consistent with the seismic structure being controlled by temperature in the presence of a hot mantle plume rising from the lower mantle. Vinnik et al.'s (2012) study was based on a larger dataset than used in the two studies of the first group (see preceding text). Vinnik et al. (2012) also argued that frequency-dependent effects may contribute to the discrepancies between the two groups.

Bowie: The presence of a narrow, low-velocity zone near the base of the transition zone below the Bowie hotspot is based on delays in seismic records of Alaskan earthquakes measured in the NW United States (Nataf and VanDecar, 1993). The delays are consistent with a zone of 150 km diameter and an excess temperature of 200 K.

Hawaii: Early tomography studies of the Hawaiian hotspot revealed the presence of anomalously slow seismic velocities in the upper mantle beneath the Hawaiian hotspot swell (Laske et al., 2007) and beneath the Hawaiian Islands (Priestley and Tilmann, 1999; Tilmann et al., 2001; Wolfe et al., 2002). Additional evidence for an upper mantle melt anomaly was provided by a seafloor magnetotelluric study (Constable and Heinson, 2004), which suggests a columnar zone of 5–10% partial melting with a radius <100 km and a depth extent of 150 km. Ritsema et al.'s (2011) global tomography model shows a broad volume of low velocities in the upper mantle beneath Hawaii that appears to connect to a low-velocity body extending through and to the base of the lower mantle (**Figure 13(g)**).

More recently, the Hawaiian Plume-Lithosphere Undersea Melt Experiment (PLUME) (Wolfe et al., 2009) involved one of the largest deployments of ocean bottom and land seismometers to image the seismic structure of the upper mantle and shallow lower mantle beneath the Hawaiian hotspot. Rayleigh wave tomography from the PLUME project reveals low velocities within the lithosphere at depths of 10–60 km beneath the island chain and the surrounding hotspot swell (Laske et al., 2011). Separate tomographic inversions of Rayleigh waves (Laske et al., 2011), S-waves (Wolfe et al., 2009), and P-waves (Wolfe et al., 2011) all image a volume of relatively low seismic velocities in the upper mantle underlying the Hawaiian swell (**Figure 16**). The volume is asymmetrical and irregular in shape. Lower seismic velocities SW of the island of Hawaii than NE correlate with an asymmetrical bathymetric swell, which is shallower to the SW than the NE (Crosby and McKenzie, 2009; Wessel, 1993), and with variations in the depth of seismic converters constrained by receiver functions (Rychert et al., 2013). These observations indicate that the Hawaiian swell is supported by hot and buoyant mantle interpreted to be plume material ponding beneath the lithosphere (Olson, 1990; Sleep, 1990) like a 'pancake' (**Section 7.10.3.4**). This pancake-shaped feature is asymmetrical about the axis of the volcano chain and appears to be hotter in – or to extend further to – the SW than the NE. In addition, all three tomography studies show that the low-velocity body is surrounded by seismically fast material, which is interpreted to be relatively cool, sublithospheric material that is sinking like a curtain around the plume pancake. The asymmetry as well as short-wavelength variability of the Hawaiian plume pancake is not

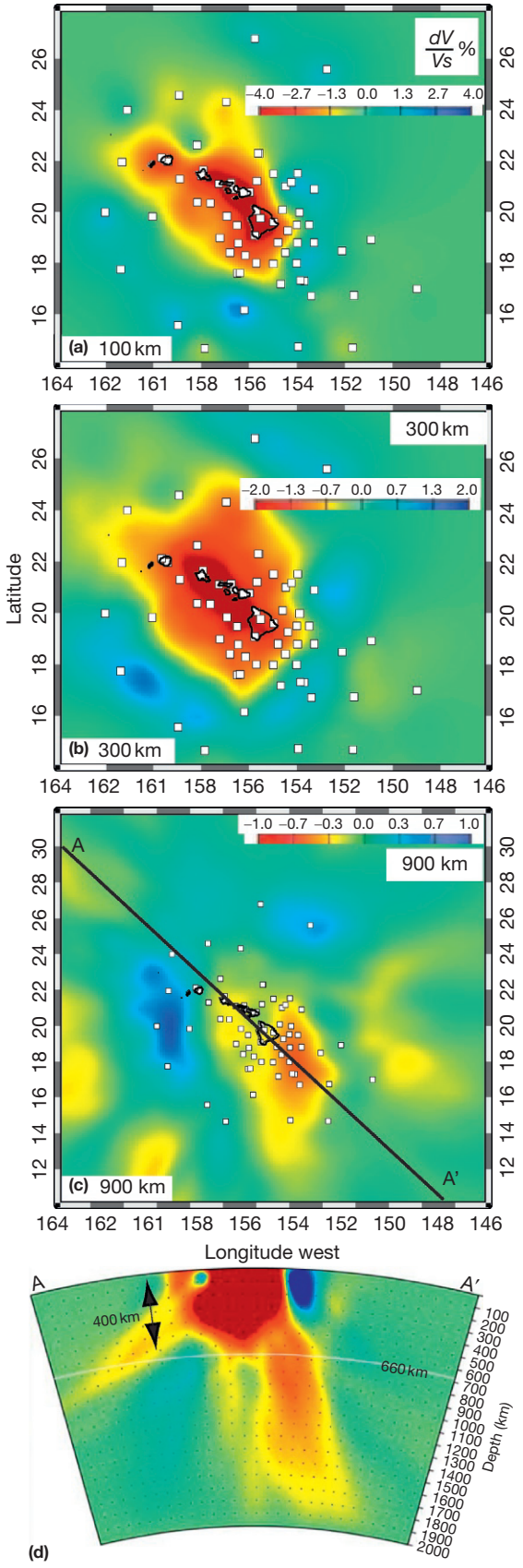


Figure 16 The Hawaiian PLUME S-wave tomography model (Wolfe et al., 2009). Horizontal cross sections at depths of (a) 100 km, (b) 300 km, and (c) 900 km of velocity variations displayed as colors with percentages labeled on the color bars. Seismometer locations are denoted with white boxes. (d) Vertical cross section along the diagonal line in and with color scale as in (c).

predicted by models of steady-state plume–plate interaction (Ribe and Christensen, 1994, 1999).

Another surprising finding of the body-wave studies (Wolfe et al., 2009, 2011) is that the low velocities underlying the Hawaiian swell are present from the base of the lithosphere all the way down to the transition zone (**Figure 16**). If these low velocities indeed identify ponding plume material, the inferred plume pancake is several times thicker than predicted by the classical models of thermally buoyant mantle plumes (e.g., Ribe and Christensen, 1999; van Hunen and Zhong, 2003). A possible explanation is that the Hawaiian plume contains compositionally dense mafic materials that reduce the net buoyancy of plume material to the degree that the rise of the plume temporarily stalls at the depths of 300–410 km, forming a 'deep eclogite pool' (Ballmer et al., 2013b; see **Section 7.10.3.3.3**). This explanation is supported by joint surface-wave and body-wave tomography (Cheng et al., in press) as well as independent constraints from receiver functions (Huckfeldt et al., 2013).

Below the upper mantle, the tomography studies image a vertically elongated body of low velocities near the southeast end of island chain that extends into the lower mantle (**Figure 16**; Wolfe et al., 2009; 2011). This feature is interpreted to be the stem of the Hawaiian mantle plume with an estimated thermal anomaly of 250–300 °C, confirming prior discoveries of transition zone thinning by 40–50 km beneath Hawaii (Li et al., 2000, 2004; Collins et al., 2002). A still deeper origin is suggested by a tomographic study that incorporates core phases (Lei and Zhao, 2006). Compelling evidence for an origin at the CMB is the imaging of an ultralow-velocity zone (ULVZ) at the base of the mantle beneath Hawaii having one of the largest known velocity reductions (~20%) yet documented (Cottaar and Romanowicz, 2012).

The Galápagos: The Galápagos hotspot is part of a broad region in the Nazca–Cocos basin with significantly reduced long-period Love and Rayleigh wave speeds (Heintz et al., 2005; Vdovin et al., 1999). The mantle transition zone in the region has a similar thickness as that of the rest of the Pacific Basin except for a narrow region of ~100 km in radius slightly to the west of the Galápagos archipelago, where it is thinned by ~18 km (Hooft et al., 2003). This amount of thinning suggests an excess temperature of ~130 K. A regional body-wave tomography study detected a low-velocity feature of comparable dimension, extending from above the transition zone into the shallow upper mantle (Toomey et al., 2001). Rayleigh wave tomography (Villagomez et al., 2007) shows laterally averaged S-wave velocities at depths 75–150 km that are 0.05–0.2 km s^{-1} lower than that beneath normal Pacific mantle with comparable seafloor ages. This velocity anomaly is consistent with excess temperatures of 30–150 °C, plus ~0.5% melt. The low-velocity material appears to trend diagonally from its greatest depth beneath the Galápagos archipelago upward and to the NE, interpreted to reflect the flow of plume material from the plume stem toward the Galápagos Spreading Center. Ritsema et al.'s (2011) global tomography

model suggests that the upper mantle anomaly connects to a broader plume-like feature extending all the way down to the CMB (**Figure 13(h)**).

To examine seismic anisotropy in the mantle, shear-wave splitting was also studied (Fontaine et al., 2005). At the western edge of the archipelago, splitting of up to 1 s with fast polarizations directed E–W is consistent with LPO and mantle deformation being dominated by the eastward motion of the Cocos Plate. Directly beneath the archipelago and within the upper mantle seismic anomaly, however, there is no clear splitting, which suggests that melt or complex flow beneath the hotspot superimposes plate motion-derived anisotropy.

Yellowstone: Our understanding of the mantle seismic structure in the area of the Yellowstone hotspot track has evolved dramatically as the data coverage has increased. Early tomographic studies revealed a complex velocity structure in the upper mantle beneath the SRP (which marks the hotspot track), interpreted to represent compositional heterogeneity associated with melting, without a deep-seated mantle plume (Saltzer and Humphreys, 1997). A subsequent study using data from local seismic networks identified low velocities in the upper mantle beneath the SRP, suggesting the presence of a plume-like feature extending as deep as 400 km to the west of Yellowstone (Waite et al., 2006; Yuan and Dueker, 2005).

The strongest constraints on mantle seismic structure near Yellowstone have been derived from data collected with Earthscope's USArray (http://www.earthscope.org) with dense seismic coverage over the whole area of the western United States. Such a dense coverage provides the opportunity to image the structure beneath Yellowstone at much higher resolution and to greater depths than beneath any other hotspot on Earth. The aperture of the seismic array provides good resolution down to depths of ~1000 km or more. Several groups produced P- and S-wave tomography models of the mantle beneath the western United States (Darold and Humphreys, 2013; Fouch, 2012; James et al., 2011; Obrebski et al., 2010; Schmandt et al., 2012; Sigloch, 2011; Tian and Zhao, 2012). The images differ in detail but common elements related to the Yellowstone hotspot include low velocities in the shallowest 300 km of the upper mantle below the SRP and a relatively narrow (<200 km wide) low-velocity feature beneath Yellowstone that extends continuously through the upper mantle and connects to a broader (~300–500 km across) low-velocity zone in the shallow lower mantle (**Figure 17**). These studies suggest that the deep low-velocity anomaly is irregular in shape, with variable amplitude, and surrounded by seismically fast material without a clear extension to depths greater than 900 km depth.

Earlier receiver function studies of the mantle transition zone found a depression of the 410 km discontinuity near Yellowstone but did not image an upward deflection of the 660 km discontinuity, as expected for the effect of high temperatures on phase changes in the olivine system (Fee and Dueker, 2004; Yuan and Dueker, 2005). The most recent receiver function study using data from the USArray, however, shows that the 660 km discontinuity shoals by 12–18 km over an area ~200 km wide and centered on the low-velocity feature (Schmandt et al., 2012). In contrast, the 410 km discontinuity appears to be only marginally perturbed (<5 km) (**Figure 17(b)**).

Interpretation of the seismic structures beneath Yellowstone remains controversial. On the one hand, there is agreement concerning the shallower structure. The seismically slow material just beneath the SRP is interpreted to be a zone of partial melting and to be bound to the NW and from below by the seismically fast remnants of the subducted Farallon Plate and to the SE by foundering lithospheric material. On the other hand, there is disagreement concerning the deeper structure, with one group (James et al., 2011) dismissing a plume origin for the seismic and volcanic anomaly altogether (see **Section 7.10.3.7.2**).

Despite these controversies, the good resolution of the seismic structure well into the lower mantle beneath both Yellowstone and Hawaii motivates a brief comparison between the two. Common features involve a horizontal low-velocity anomaly beneath the hotspot tracks that extends in the direction of plate motion, as well as a narrow vertical anomaly near the leading edge of the hotspots that protrudes into the lower mantle. The simplest explanation for these two common features is a plume rising from the lower mantle and ponding beneath the lithosphere. A notable difference is that the deep anomaly beneath Yellowstone is not clearly continuous with depth compared to the inferred continuity for that beneath Hawaii. Another difference concerns the depth extent of the horizontal upper mantle anomaly, which is completely contained in the shallowest 200–300 km beneath Yellowstone's hotspot track (SRP), whereas it extends through most of the upper mantle (Wolfe et al., 2009) or even splits up into two separate layers (Cheng et al., in press) beneath the Hawaiian Islands.

Eifel: The Eifel region in western Germany is characterized by numerous but small volcanic eruptions with contemporaneous uplift of 250 m in the last 1 My. Tomographic imaging indicates a mantle low-velocity anomaly extending to depths of at least 200 km (Passier and Snieder, 1996; Pilidou et al., 2005). Inversions using a high-resolution local array study indicate a fairly narrow (100 km) P-wave anomaly of 2% that possibly extends to 400 km (Keyser et al., 2002; Ritter et al., 2001). The connection with the deep mantle is unclear but has been suggested to include the low-velocity structure that is imaged in the lower mantle below central Europe (Goes et al., 1999). Shear-wave splitting measurements show the largest split times for S-waves polarized in the direction of absolute plate motion, but the pattern is overprinted by complex orientations, suggestive of parabolic mantle flow around a plume stem with the overlying plate moving to the SW (Walker et al., 2005).

Cameroon Volcanic Line: Body-wave tomographic studies using global data and those obtained from the recent Cameroon Broadband Seismic Experiment show that the Cameroon volcanic line (CVL) is underlain by a fingerlike low-velocity anomaly that extends from the Gulf of Guinea in the SW to the Chad–Cameroon border in the NE (e.g., Reusch et al., 2010; Ritsema et al., 2011). The anomaly ranges from ~100 to ~400 km depth with the depth of peak amplitudes shoaling toward the SW, and is juxtaposed by a high-velocity anomaly associated with the Congo Craton (and subcratonic mantle) toward the SE (Reusch et al., 2010). The peak-to-peak difference between the two anomalies is ~5%, corresponding to a temperature contrast of >280 °C (Reusch et al., 2010).

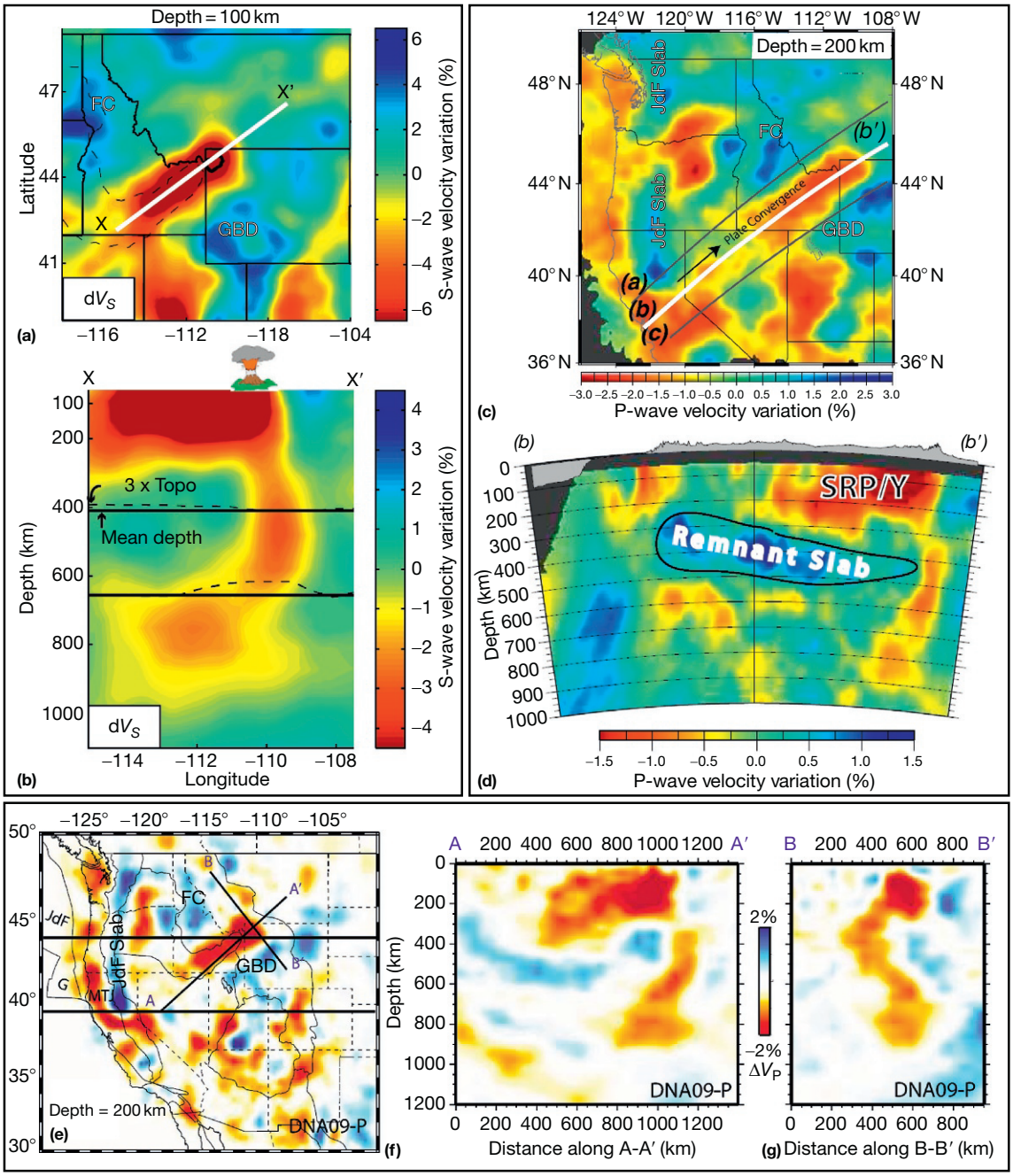


Figure 17 Tomography results for (a) and (b) S-waves by Schmandt et al. (2012), (c) and (d) P-waves by James et al. (2011), and (e)–(g) P-waves by Obrebski et al. (2010). For each tomography model, a horizontal cross section at a given depth (labeled) is shown, along with a SW–NE, vertical cross section along the Snake River Plain (SRP) and through the Yellowstone (Y) hotspot (locations labeled on maps). For one model, a NW–SE vertical cross section is also shown (g). GBD = Great Basin (lithospheric) drip (West et al., 2009); FC = Farallon 'curtain' (Darold and Humphreys, 2013); JdF, Juan de Fuca; G, Gorda Plate; 'MTJ', Mendocino triple junction. (b) Depth variations (dashed curves) of seismic discontinuities indicate a shoaling at 660 km and no change at 410 km beneath Yellowstone compared to the mean depths of the discontinuities (solid).

This contrast may suggest anomalously hot mantle beneath the CVL, anomalously cool mantle beneath the Congo Craton (e.g., due to downwelling; King and Ritsema, 2000), or both. A rather moderate thermal anomaly beneath the CVL itself is consistent with receiver function constraints on the ratio of P- to S-wave speed (VP/VS), which indicate the presence of only very small fractions of melt in the asthenosphere (Gallacher and Bastow, 2012).

Shear-wave splitting measurements display fast directions dominantly parallel or subparallel to the CVL (Koch et al., 2012). These directions are consistent with a strong component of local northeastward asthenospheric flow, as is

predicted by numerical simulations to result from large-scale mantle convection patterns (Conrad and Behn, 2010; Forte et al., 2010). A weaker component of flow perpendicular to the CVL, perhaps related to small-scale convection along the edge of the Congo Craton (cf. King and Ritsema, 2000), may contribute to subparallel (i.e., dominantly N–S) splitting directions measured by Koch et al. (2012) at a subset of stations. Such edge-driven convection (see **Section 7.10.3.3.5**) may also provide a good explanation for volcanism itself (King and Ritsema, 2000). An alternative plume origin for volcanism is ruled out by nonanomalous transition zone thicknesses (Reusch et al., 2011) and the lack of hotspot-like age–distance relations (**Figure 6**, inset).

East Africa: The hotspots in East Africa overlie the northeastern flank of the African LLSVP (**Figure 13(e)**). The African LLSVP is rooted at the CMB beneath southern Africa (Helmberger et al., 2000). It is a large low-velocity body (~4000 km by ~2000 km) that tilts up and to the northeast beneath East Africa (Forte et al., 2010; Ritsema et al., 1999) (**Figure 13(i)**). The relationship of the African LLSVP to volcanism and rifting in East Africa remains controversial. One suggestion is that volcanism is not attributed to the LLSVP at all, but instead to small-scale, sublithospheric convection along the edges of the African Craton (King and Ritsema, 2000). Another idea attributes the numerous volcanic centers to one or more mantle plumes of ~100 km in width rising from the top of the LLSVP to the base of the lithosphere (e.g., Chang and Van der Lee, 2011; Ebinger and Sleep, 1998; Montelli et al., 2006). A third concept attributes volcanism directly to the rise into the shallow upper mantle and decompression melting of a 'superplume' (cf. Davaille, 1999) as a whole, which would be imaged as an extension of the LLSVP itself (Bastow et al., 2005; Benoit et al., 2006; Park and Nyblade, 2006).

With the arrival of data from permanent and temporary seismic arrays of the AfricaArray (http://www.africaarray.psu.edu/), a general consensus is forming among the most recent studies (Hansen and Nyblade, 2013; Hansen et al., 2012). S-wave (Nyblade, 2011) and surface-wave (Adams et al., 2012) tomography reveal that the shallow upper mantle in East Africa from just south of the Tanzania Craton to NE of the Ethiopian Rift is seismically slow (**Figure 18**). Associated with this low-velocity anomaly, the 410 km seismic discontinuity is 30–40 km deeper than normal suggesting excess temperatures of >250 °C (Cornwell et al., 2011; Huerta et al., 2009). The seismically slow material in the upper mantle is therefore too thick, geographically broad, and pervasive to be explained by the shallow process of edge-driven convection or by separate, smaller-scale mantle plumes. The observed seismic structure appears to be most consistent with the penetration of a large and broad (super)plume into the upper mantle where it causes decompression melting. Fast polarization directions of shear-wave splitting that are dominantly oriented NE–SW have been interpreted to reflect LPO due to dominantly

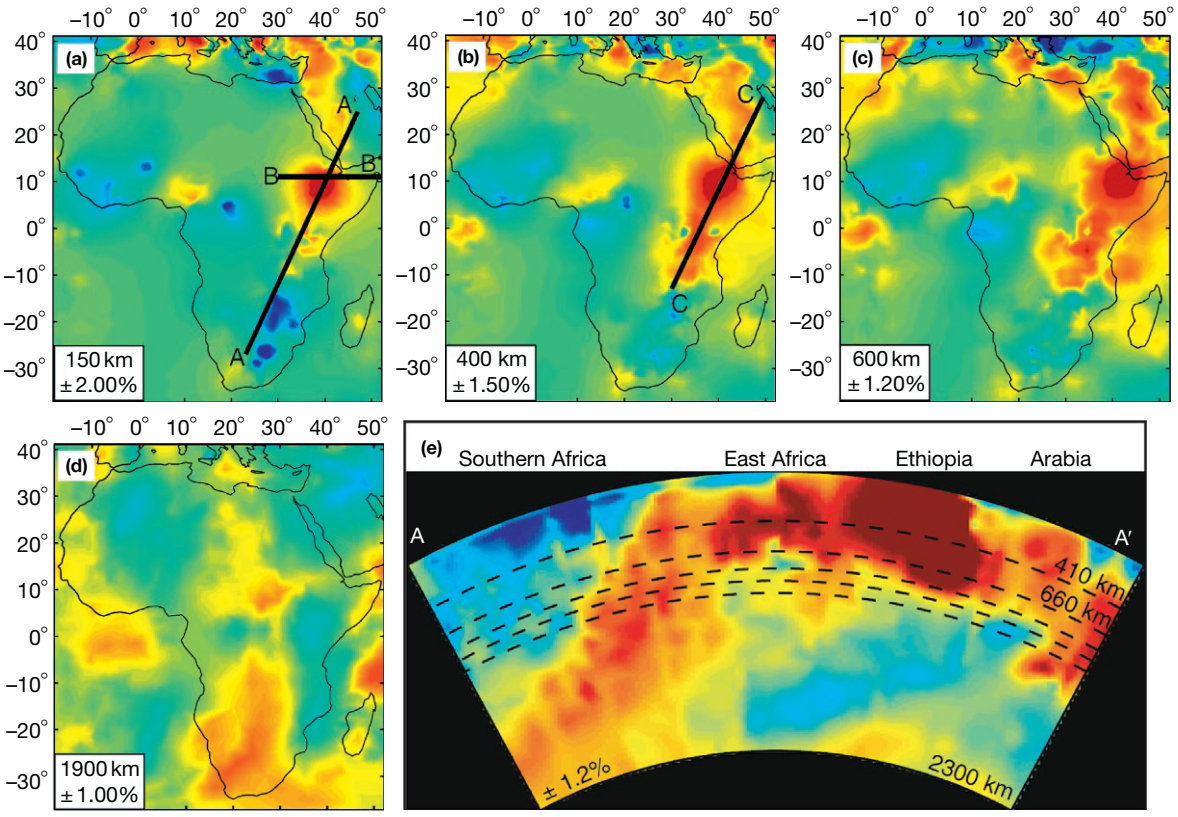


Figure 18 Tomography images of the mantle beneath Africa from P-wave inversions of Hansen et al. (2012). (a)–(d) Horizontal cross sections of P-wave velocity perturbations at the labeled depths. (e) Vertical cross section along the profile A–A′ as marked in (a). The ± values indicate the bounds of the color scales in each corresponding panel; blue is fast, while red is slow. Dashed lines in (e) mark depths of 410, 660, 800, 900, and 1000 km.

northeastward flow of the superplume in the upper mantle (Bagley and Nyblade, 2013), an interpretation that is supported by geodynamic models (Forte et al., 2010).

Potential challenges for a superplume origin arise from a recent S-wave tomography study that images vertically elongate low-velocity bodies beneath Afar and Kenya that protrude downward below the low-velocity upper mantle to at least midmantle (~1400 km) depths (Chang and Van der Lee, 2011). Also, the 660 km seismic discontinuity beneath East Africa is regionally abnormally deep (Cornwell et al., 2011; Huerta et al., 2009). In the context of the superplume hypothesis, this observation would require the presence of eclogitic materials to activate the majorite–perovskite phase change at ~720 km depth (Cornwell et al., 2011; Huerta et al., 2009). A SW–NE shoaling of the discontinuity from northern Ethiopia to Afar would then reflect a decrease in the garnet concentration (Cornwell et al., 2011).

South Pacific Superswell: Approximately antipodal to East Africa is a cluster of hotspots in the area of the South Pacific Superswell (McNutt and Fischer, 1987), which overlies the Earth's other LLSVP in the lower mantle (e.g., Breger et al., 2001; Ni et al., 2005; Trampert et al., 2004; van der Hilst and Karason, 1999). Ritsema et al.'s (2011) S40RTS model shows that the whole area of the South Pacific Superswell is underlain by low seismic velocities in both the lower and the upper mantles (**Figure 13(j)**).

A recent deployment of temporary seismometers on the seafloor and a number of islands enabled a regional tomographic study of the middle and upper mantle beneath the South Pacific Superswell. The P-wave and S-wave tomography results reveal that the area is underlain by multiple low-velocity regions in the depth range from 400 to 1600 km (Tanaka et al., 2009). Two low-velocity regions at 1200 km depth are centered beneath the Society and Pitcairn hotspots, while two at 800 km depth are located beneath the Tuamotu and Austral Islands. There are also relatively short-wavelength (hundreds of km across) variations at 400 km depth. A prior study of underside reflections of S-waves at the 410 and 660 km discontinuities found the mantle transition zone to have normal thicknesses, except in a ~500 km wide area beneath the Society hotspot (Niu et al., 2002a).

The incorporation of Rayleigh wave tomography highlights the presence of short-wavelength variability in the upper mantle, showing that the Society, Pitcairn, Macdonald, Marquesas, and Arago hotspots are underlain by low velocities in the depth range of 60–180 km (Suetsugu et al., 2009). Only the low-velocity bodies beneath Society and Macdonald appear to connect to the low-velocity anomalies in the lower mantle. The shallow anomalies beneath the other hotspots appear to be restricted to the upper mantle. Suetsugu et al. (2009) interpreted this structure as reflecting the presence of a number of small-scale (hundreds of km across), plume-like upwellings emanating from the south Pacific LLSVP, some of which having risen above and become separated from the LLSVP. This conclusion is in contrast to the conclusions for the African LLSVP rising as a whole superplume into the shallow upper mantle. These contrasting results lead to questions about the dynamics of LLSVPs. For example, are the contrasting dynamics related to internal forces (differences in composition and temperature) or instead driven by external ones (ambient mantle flow shaping the LLSVPs; McNamara and Zhong, 2005; Steinberger and Torsvik, 2012)?

7.10.2.6 Summary of Observations

Intraplate volcanic activity can be mainly grouped in three categories: long-lived age-progressive volcanism, short-lived age-progressive volcanism, and volcanism with complex age patterns (Courtillot et al., 2003). Long-lived (>50 My) age-progressive volcanism is documented for at least 13 hotspots. At present day, these hotspots define a global kinematic reference frame that is deforming slower than average, present-day plate velocities. Over geologic time, however, the motion between the Indo-Atlantic hotspots, the Pacific hotspots, and Iceland has been significant. Short-lived (≤24 My) age-progressive volcanism is documented along at least ten volcano chains. The directions and rates of age progression in the short-lived chains suggest relative motion between these hotspots, even on the same plate. Finally, the available dates of rocks from a number of volcano groups (e.g., Line and Cook–Austral Islands and Cameroon volcanic line) do not show evidence for simple age–distance relations but instead suggest these groups to have formed by episodic volcanism over tens of millions of years with coeval volcanism over large distances.

Hotspots are commonly associated with topographic anomalies (i.e., swells). These swells are usually centered by the most active volcanoes, span distances of hundreds of km to >1000 km, and rise hundreds of meters above the surrounding seafloor. Hotspot swells usually diminish with time or distance from the center of active volcanism and exclusively support volcanoes no older than ~50 My.

LIPs display a huge range in magmatic volumes and durations and have been related to climate change and mass extinction events. They represent the largest outpourings of magma from as large as 50 Mkm^3 (OJP) to <2 Mkm^3 (CRB) (see **Figure 10**). This voluminous magmatism can occur in a dramatic short bursts, lasting 1–2 My (CAMP, CRB, EM, OJP, and SIB), or can be prolonged over tens of My (e.g., CHON, CBN, KER, NAVP, and SHA). Main eruptive products are tholeiitic basalts that, on the continents, typically are transported through radiating dike swarms. High-MgO basalts or picrites are found in a number of provinces (NAVP, OJP, and CAR) indicating high degrees of mantle melting. Smaller rhyolitic eruptions in continental flood basalts (KAR, PAR, CHON, and EARS) instead suggest secondary melting of the continental crust. Dynamic topographic uplift (analogous to hotspot swells) is evident around the main eruptive stages of some LIPs (EME, NAVP, SIB, and KER) but may not have occurred at others (OJP and SHA). While an appreciable amount of the geologic record is lost to subduction, about six active hotspots are confidently linked to LIPs. Most of these hotspots are currently located at or very close to an oceanic spreading center with the Kerguelen hotspot being a notable exception. Flood basalt volcanism itself is also typically associated with rifting, either between continents (PAR, KAR, CAMP, NAVP, DEC, and MAD) or at MORs (OJP–MAN–HIK, KER, and SHA–HES). These links between LIPs and plate tectonics compel substantial revisions or alternatives to the hypothesis of an isolated head of a starting mantle plume as the sole origin of LIPs.

Basalts from hotspots and other sites of intraplate volcanism, commonly referred to as OIB, are geochemically distinct from MORB. Both isotope ratios and major elements display a much greater variability among OIB than among MORB. Generally, OIBs are more mafic (i.e., lower SiO_2 content) and alkalic (higher $Na_2O + K_2O$) than MORB. For the most part, they also differ in at least one of three commonly analyzed isotope ratios: $^{87}Sr/^{86}Sr$, $^{206}Pb/^{204}Pb$, and $^{3}He/^{4}He$. Exceptions are the Bowie–Kodiak, Cobb, and Caroline chains, which show MORB-like $^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$ compositions (but lack $^{3}He/^{4}He$ data). In terms of major elements, the most productive hotspots (Hawaii, Iceland, and the Galápagos) overlap most heavily with MORB.

Hotspots are typically associated with anomalously low seismic-wave speeds below the lithosphere and in the upper mantle. Transition zone thicknesses are often anomalously thin by tens of km. The previously mentioned findings are consistent with elevated mantle temperature (by 150–200 K) and with the abundance of partial melt in the shallow upper mantle. Improved understanding of mineral physics at appropriately high pressure and temperature is needed to better constrain the magnitude of the putative temperature anomalies and to quantify the potential contribution of compositional heterogeneity. Especially with recent regional studies and the availability of EarthScope data, we have seen increasing evidence for a lower mantle origin for a selected number of hotspots.

The key characteristics described earlier provide important constraints to test dynamic mechanisms proposed for the origin of melting anomalies. It seems unlikely that a single overarching mechanism can account for the diversity seen in the geologic record. In the next section, we discuss the implications and predictions of each of these mechanisms.

7.10.3 Dynamic Mechanisms and Their Implications

This section reviews the mechanisms proposed to generate intraplate volcanism. We begin with a summary of methods that are used to quantitatively explore dynamic mechanisms (Section 7.10.3.1). We then discuss the geodynamic processes for generating mantle upwelling (Section 7.10.3.3) and related magmatism (Section 7.10.3.2). Candidate mechanisms to compensate hotspot swells and to sustain flood basalt volcanism are explored in Sections 7.10.3.4 and 7.10.3.5, respectively. In the context of the discussed geodynamic processes, we continue revisiting the geochemical and petrologic signatures of hotspot lavas and their distinctions from MORB (Section 7.10.3.6). Finally, we discuss these processes in the broader context of mantle convection and plate tectonics (Section 7.10.3.7).

7.10.3.1 Methods

The origin and evolution of hotspots and melting anomalies can be constrained by systematically comparing predictions of geodynamic models with observations. Such models simulate the transport of energy, mass, and momentum in the solid mantle that may have local patches of partial melt. They indeed provide a means to directly test conceptual ideas against the basic laws of physics and to delineate the conditions under which a proposed geodynamic mechanism can operate.

Key aspects of material and energy transport in the mantle can be described mathematically and solved by analytic or numerical approaches or can be studied in laboratory experiments using analogue materials. Analytic approaches provide scaling laws that reveal relationships between phenomena and key parameters for simple or simplified problems (see also Chapter 7.04). Numerical simulations can address much more complex problems. Their accuracy, however, may be limited by the discretization of the differential equations and the parameterizations made (see also Chapter 7.05). Laboratory experiments (see also Chapter 7.03) involving analogue materials such as sucrose syrup and silicon putty instead provide natural experiments that do not suffer from numerical discretization but require significant extrapolation over length scales and timescales. For simplified cases, such as the ascent of plumes through a nearly isoviscous or Newtonian fluid, benchmarks have revealed good agreement of numerical and analogue simulations (e.g., van Keken's (1997) numerical benchmark of the classic Griffiths and Campbell (1990) laboratory experiments; see also Vatteville et al. (2009), van Keken et al. (2013)). It is nevertheless essential to resort to computational modeling to simulate some of the complexities that arise with the features of the Earth's mantle such as compressibility, phase transitions, melting, and realistic rheology dependent on grain size, pressure, and stress (see also Chapter 7.02).

7.10.3.2 Generating Magmas for Volcanism

Understanding the causes for melt generation is essential for studying the origin of hotspots and intraplate volcanism. Partial melting of rocks can conceptually occur due to (1) a change in temperature, (2) a closed-system decompression, and (3) an open-system change in composition. While the scenarios (1) and (3) are relevant for melting the crust (e.g., near intrusions) and the mantle wedge (i.e., at subduction zones), respectively, isentropic decompression melting (2) is the dominant process of melt generation at MORs, hotspots, and other melting anomalies.

For decompression melting, the total volumetric rate of melt generation is approximately proportional to the rate of decompression (or mantle upwelling) and the melt productivity at constant entropy $(-\partial F/\partial P)_S$ integrated over the volume V of the melting zone:

$$Q_M = \frac{\rho_c}{\rho_c}\int_V \left(-\frac{\partial F}{\partial P}\right)_S \left(-\frac{\mathrm{D}P}{\mathrm{D}T}\right) \mathrm{d}V \qquad [1]$$

where ρ_m is mantle density and ρ_c is igneous crustal density. $(-\partial F/\partial P)_S$ is generally positive above the solidus and zero otherwise. The solidus temperature and $(-\partial F/\partial P)_S$ depend on the equilibrium composition of the solid and liquid at a given pressure (e.g., Hirschmann et al., 1999; McKenzie, 1984; Phipps Morgan, 2001). Mantle magmatism thus requires some combination of excess temperature, presence of fertile or fusible material, and mantle upwelling. Higher temperatures boost melting by increasing the pressure range over which the solidus is exceeded, composition controls melting by changing both this pressure range and $(-\partial F/\partial P)_S$, and both factors may influence the rate of decompression $(-\mathrm{D}P/\mathrm{D}T)$ through their effects on mantle buoyancy and upwelling rate.

7.10.3.2.1 Temperature

Excess melting sustained by elevated temperatures has been a major focus of previous studies. Mantle-source temperatures are commonly estimated from the Fe–Mg content of primary magmas and the olivine phenocrysts with which they equilibrate. One group suggests that the mantle is no hotter beneath Hawaii than beneath many MORs (Falloon et al., 2007; Green et al., 2001). Other groups, however, suggest elevated temperatures of, for example, 100–300 °C beneath Hawaii, 50–100 °C beneath Iceland or the Azores, and ~150 °C beneath Afar (Beier et al., 2012; Herzberg, 2004a; Herzberg and Gazel, 2009; Herzberg et al., 2007; Lee et al., 2009; Putirka, 2005; Putirka et al., 2007, 2011; Rooney et al., 2012). Most, if not all, lavas sampled at hotspots have evolved to varying degrees after leaving the mantle source, and it is therefore difficult to estimate the parental liquids' MgO content. For example, Putirka (2005) argued that the lower MgO contents derived by Green et al. (2001) for Hawaii could lead to an underestimate of temperature. Restricting the analyses to melt inclusions would mitigate this uncertainty, but such an approach is intricate.

Complementary methods rely on peridotite melting models that invert from or forward model a whole suite of major-element and/or minor-element concentrations, sometimes including crustal thicknesses as an additional constraint. Beneath Iceland, estimates for excess mantle temperatures based on such methods range from ~100 to ~250 °C (e.g., Herzberg and O'Hara, 2002; Klein and Langmuir, 1987; Langmuir et al., 1992; Maclennan et al., 2001; McKenzie and Bickle, 1988; Presnall et al., 2002; Shen and Forsyth, 1995; White and McKenzie, 1995). Excess temperatures are further estimated at 200–300 °C beneath Hawaii, based on a combination of geodynamic and melting models (Ribe and Christensen, 1999; Watson and McKenzie, 1991), and at ~100 °C beneath Afar, based on incompatible trace-element forward modeling (Ferguson et al., 2013). Along these lines, most of the studies appear to converge to high thermal anomalies beneath Hawaii and moderate thermal anomalies at other hotspots, but discrepancies between methods remain large.

7.10.3.2.2 Composition

An additional source of uncertainty for the previously mentioned temperature estimates is the composition of the mantle source. Volatile species such as H_2O and CO_2 can dramatically reduce solidus temperatures even in small proportions, as are present in the MORB source (Asimow and Langmuir, 2003; Dasgupta and Hirschmann, 2006; Dasgupta et al., 2007; Hirschmann, 2010). While such small concentrations of volatiles are not likely to increase the total extent of melting significantly, they can enhance the amount of melt produced for a given temperature by expanding the volume of the melting zone. As the mantle beneath hotspots is thought to be more volatile-rich than beneath MORs, temperature estimates based on dry peridotite may be too high. For example, excess temperatures beneath the hotspot-influenced Galápagos Spreading Center may be reduced from ~50 °C for anhydrous melting models to <40 °C when water is considered (Asimow and Langmuir, 2003; Cushman et al., 2004). Similarly, estimates for the mantle excess temperature beneath the Azores have been revised from ~75 to ~55 °C (Asimow and Langmuir, 2003). Hydrous melting models have yet to be explored in detail for the larger Iceland and Hawaii hotspots.

The mantle beneath hotspots may also contain more fusible, mafic lithologies, such as those generated by the recycling of subducted oceanic crust. The presence of more fusible or 'fertile' mantle has been suggested for hotspots such as Hawaii (Hauri, 1996; Herzberg, 2006; Sobolev et al., 2005; Takahashi, 2002), Iceland (Korenaga and Kelemen, 2000), the Columbia River basalts (Takahashi et al., 1998), the Galápagos (Sallares et al., 2005), the Canaries (Day et al., 2009; Gurenko et al., 2009; Herzberg, 2011), and others (Hofmann, 1997). Mafic lithologies tend to have a lower solidus and much greater isentropic melt productivity $(-\partial F/\partial P)_S$ than peridotite (Pertermann and Hirschmann, 2003; Spandler et al., 2008; Yasuda and Fujii, 1994) and therefore require significantly lower temperatures to produce the same volume of magma compared to peridotite.

Some have argued that such fertile mantle melting could generate large melting anomalies with very small or even no excess temperatures (Korenaga, 2005). An important difficulty with this hypothesis is that mafic materials will tend to form eclogite, which is significantly denser than lherzolite throughout the upper mantle (Aoki and Takahashi, 2004; Hirose et al., 1999; Irifune et al., 1986). To undergo decompression melting, this material would therefore have to be supported by rapid ambient mantle upwelling. Korenaga proposed that rapid upwelling driven by shallow small-scale convection (cf. **Section 7.10.3.3.5**) or fast seafloor spreading could entrain eclogite upward from the uppermost lower mantle (Korenaga, 2004, 2005), where eclogite becomes positively buoyant (Hirose et al., 1999) and hence perhaps accumulates. More recent experiments, however, suggest that the excess density of quartz-normative eclogites is actually near doubled in the depth range of 300–410 km (Aoki and Takahashi, 2004), which would make it difficult for plate-driven mantle flow to entrain significant amounts of eclogite into the asthenosphere. A more complete understanding of the properties and phase relations of different lithologies at a range of mantle pressures and temperatures is needed to test the ability of fertile materials to give rise to mantle melting without thermal anomalies.

7.10.3.2.3 Mantle upwelling

The final major factor that can lead to melting anomalies is enhanced mantle upwelling (which determines the decompression rate, $-DP/Dt$). Not only thermal buoyancy but also compositional buoyancy may fuel upwelling. However, compositionally lighter materials such as those with less iron and garnet than the ambient mantle, perhaps due to prior melting (Jordan, 1979; Oxburgh and Parmentier, 1977), are typically less fertile than dense, undepleted materials.

Behaving in complementary fashion to fertile mantle, depleted mantle must be light enough such that the associated increase in upwelling ($-DP/Dt$) overcomes the reduction in fusibility $(-\partial F/\partial P)_S$ (cf. eqn [1]).

7.10.3.3 Dynamics of Mantle Upwellings

Geodynamic modeling studies have explored a wide range of processes that can create and influence mantle upwellings. Upwellings may occur in columnar plumes or diapirs. The

style of ascent of such upwellings can be greatly affected by variable rheology, thermal expansivity, and conductivity. Compositional heterogeneity can further complicate ascent styles if entrained by plumes. Alternatively, upwelling may be passively driven, for example, by downwellings that remove the cool and negatively buoyant base of the lithosphere or by vertical motion of the whole lithosphere itself. Another form of vertical motion in the mantle, shear-driven upwelling, is independent of density variations as an energy source, but is rather induced by shearing of a rheologically variable mantle.

7.10.3.3.1 Thermal boundary layer instabilities

Density variations near thermal boundary layers can become convectively unstable and lead to plume-like upwelling. In the simplest case of a thermally stratified mantle, the energy source comes from a density inversion with cool rocks overlying hot rocks. Upwelling thermal instabilities develop from small anomalies within such a hot thermal boundary layer as long as they grow faster by advection than they dissipate by thermal diffusion. In the Earth's mantle, the core–mantle boundary (CMB) is the main candidate for a plume-generating thermal boundary layer (e.g., Boehler, 2000), but other boundary layers may exist at locations where sharp transitions in material properties or composition occur, such as the bottom of the transition zone at 670 km depth or the top of a proposed thermochemical layer in the deep mantle.

In its simplest form, the growth of an upwelling from a hot boundary layer can be approximated as a Rayleigh–Bénard instability with the onset time and growth rate controlled by the local (or boundary) Rayleigh number:

$$Ra_\delta = \frac{\rho g \alpha \Delta T \delta^3}{\mu \kappa} \quad [2]$$

The instability is enhanced by larger thermal expansivities α, temperature jumps across the boundary layer ΔT, or layer thicknesses δ and hampered by higher viscosities μ and effective thermal diffusivities κ. Of all these parameters, mantle viscosity μ is the least well constrained and thus μ is the largest source of uncertainty in Ra_δ. For more specifics on the governing equations for boundary layer instabilities and examples of their modeling with laboratory and numerical techniques, see Chapters 6 and 11 of Schubert et al. (2001). Analytic methods provide important insights into the rate of formation of the instability and the dependence on ambient conditions (see, e.g., Whitehead and Luther, 1975; Ribe and de Valpine, 1994). The growth of the instability to a full diapir or plume can be understood with nonlinear theory. For example, Bercovici and Kelly (1997) showed that growth may be retarded due to draining of the source layer by an intermittently stalling diapir. Experimental and numerical work confirms and expands on these predictions (e.g., Olson et al., 1988; Ribe et al., 2007). In general, most studies find that for reasonable lower mantle conditions, boundary layer instability will grow on a timescale of about 10–100 My (e.g., Christensen, 1984; Olson et al., 1987; Ribe and de Valpine, 1994; see also **Chapter 7.04**).

7.10.3.3.2 Thermal plumes

Analogue laboratory simulations of convection without a large-scale background mantle flow show that an unstable boundary layer can lead to numerous simultaneous plumes that interact with each other as they rise through the fluid (e.g., Kelly and Bercovici, 1997; Lithgow-Bertelloni et al., 2001; Manga, 1997; Olson et al., 1987; Whitehead and Luther, 1975). To study the dynamics of a single plume, it has become common to use a more narrow, point-like source of heat, which in laboratory experiments can be achieved by inserting a small patch heater at the base of the tank (e.g., Davaille and Vatteville, 2005; Kaminski and Jaupart, 2003) or alternatively by injecting hot fluid through a small hole (Griffiths and Campbell, 1990). These single-plume experiments predict the formation of a broad rounded head that leads the rising thermal instability and is trailed by a (usually) thinner columnar tail that connects to the deep source of the plume.

The morphology of the plume head and tail is controlled by the viscosity contrast between the hot plume and the ambient fluid (see Ribe et al. (2007) for a thorough review). A more viscous plume will tend to form a head with approximately the same width as the tail (a 'spout' or 'finger' morphology), whereas a lower-viscosity plume will tend to form a voluminous plume head much wider than the tail (a 'mushroom' or 'balloon' morphology). Since mantle viscosity is a strong function of temperature, the mushroom or balloon geometry should dominate thermal plumes (Kellogg and King, 1997; Olson and Singer, 1985; Richards and Griffiths, 1989; Whitehead and Luther, 1975).

The rise speed of plumes is proportional to its buoyancy and the square of its radius and inversely proportional to viscosity. Uranium-series geochemistry provides constraints for ascent rates near the base of the lithosphere; these upwelling speeds are often found to be comparable to plate velocities (Bourdon et al., 1998, 2005, 2006; Kokfelt et al., 2003), except for the much faster (i.e., by 1–2 orders of magnitude) rising Hawaiian plume (Pietruszka et al., 2001; Sims et al., 1999).

A range of instabilities, however, can induce oscillations in rise speeds and mass fluxes of thermal plumes. For example, perturbations within or above the deep thermal boundary layer can induce solitary waves that travel up the plume conduit much faster than average plume ascent rates for strongly temperature-dependent rheology (Olson and Christensen, 1986; Olson et al., 1993; Schubert and Olson, 1989; Scott et al., 1986; Whitehead and Helfrich, 1988). Other experiments have shown that a strongly tilted plume tends to break up into multiple diapiric upwellings (Olson and Singer, 1985; Richards and Griffiths, 1989; Skilbeck and Whitehead, 1978).

Such instabilities can have important implications for thermal entrainment of ambient mantle materials, through which the plume rises. In contrast to the classic work by Griffiths (1986), Richards and Griffiths (1989), and Griffiths and Campbell (1990, 1991), numerical models have revealed that thermal entrainment is generally limited to the periphery of plume tails (and even of plume heads) and hence insufficient to allow for partial melting of entrained materials beneath the hotspot (Farnetani and Hofmann, 2009; Farnetani and Richards, 1995; Farnetani et al., 2002). Solitary waves traveling rapidly and in isolation up the plume conduit moreover preclude any mixing of the hottest plume pulses with the ambient mantle (e.g., Whitehead and Helfrich, 1988). In contrast, strongly tilted plumes may be able to entrain ambient material more efficiently, perhaps even into the hot center of the

conduit (Richards and Griffiths, 1989). Understanding thermal entrainment is essential for constraining the depths of origin of geochemical signatures identified in hotspot lavas (cf. **Section 7.10.3.6**).

7.10.3.3.3 Thermochemical plumes

Thermal plumes rising from deep boundary layers have certainly guided much of our understanding of the causes for the largest hotspots. It is nevertheless likely that the density of many plumes is influenced by composition in addition to temperature. A large volume of compositionally dense but still hot and thermally buoyant material is thought to be present in the deep part of the lower mantle (Kellogg et al., 1999; McNamara and Zhong, 2005) and feed rising thermochemical plumes (e.g., Deschamps et al., 2011).

The structure of the thermochemical boundary layer at the base of the mantle can control the thermal, compositional, and rheological structure of the rising plumes. Without rapid ambient flow, materials originating from deeper versus shallower parts of the layer are expected to be entrained into the more central versus more peripheral plume conduit, respectively (Farnetani and Hofmann, 2009). If the layer is compositionally stratified, the deepest (i.e., densest albeit hottest) parts of the layer tend to remain stable and not rise into the plume (Farnetani, 1997; Lenardic and Jellinek, 2009). In addition to effects on plume composition and rheology (and therefore ascent style), such a situation may result in plume temperatures significantly smaller than CMB temperatures. These moderately hot plumes, further cooling along an adiabat that is steeper than that of the ambient mantle (Leng and Zhong, 2008), may reconcile the discrepancy between temperature anomalies at hotspots (e.g., Herzberg et al., 2007) and the much larger estimates for the temperature jump across the CMB (e.g., Boehler, 2000).

In addition, ambient mantle flow can trigger the creation of a plume and influence its thermochemical structure. For example, large-scale mantle convection tends to modulate the topography of the thermochemical boundary layer to focus plume upwellings (Steinberger and Torsvik, 2012). On a more regional scale, Tackley (2011) showed that the impingement of downgoing slabs on a thermochemical boundary layer can trigger plumes of various types (**Figure 19**). The angle of slab impingement influences the amount of mafic (and harzburgitic) materials entrained by, and the distribution of these materials within, the plume (e.g., the presence of mafic and/or refractory harzburgitic materials in the plume head). These predictions are relevant for the geochemical expression of LIP and hotspot volcanism, as well as LIP–hotspot connections (see **Chapter 7.12**).

Due to a competition between negative chemical and positive thermal buoyancy forces, entrainment of compositionally dense (e.g., mafic) materials strongly affects thermochemical-plume behavior (**Figure 20**). Plume behavior is predominantly controlled by the lateral thermal and compositional buoyancy profiles across the plume. These profiles depend on the temperature and density contrasts across the boundary layer from which they rise, the thickness of the chemical layer, and the contrast in compressibility between the two materials (Huang et al., 2010; Lin and van Keken, 2006a,b,c; Samuel and Bercovici, 2006; Tan and Gurnis, 2007). Predicted plume

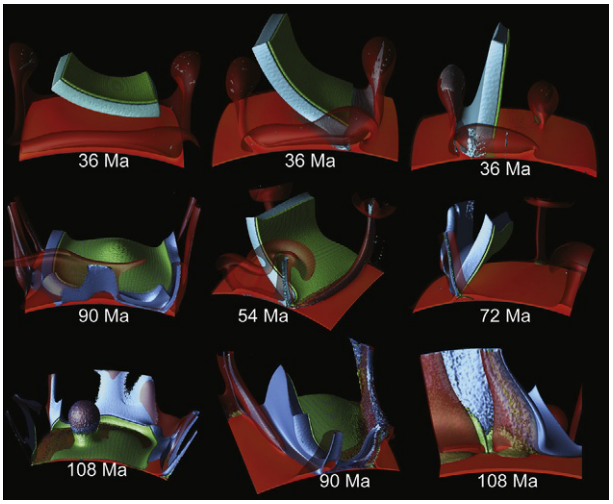


Figure 19 From Tackley (2011). Thermochemical plumes triggered by downgoing slabs that interact with a thermochemical boundary layer for three different initial slab angles (different columns). Green and blue isosurfaces denote eclogitic and harzburgitic material, respectively. The red isosurface is the 3000 K isotherm.

ascent styles for different density (**Figure 20(e)–20(l)**) and compressibility contrasts range from nearly stagnant, wide plumes in the lower mantle ('domes' or 'superplumes') to fast episodic pulsations traveling up the plume conduit (e.g., Davaille, 1999; Kumagai et al., 2008; Lin and van Keken, 2005). For somewhat lower density contrasts, the dense layer may get entrained into the plume and then cause the ascent of the plume to stall in the lower mantle (Samuel and Bercovici, 2006) or to oscillate for a range of conditions (Davaille, 1999; Lin and van Keken, 2006b,c; Huang, 2008). The large range in predicted morphologies combined with the potential for stalling or sinking of thermochemical plumes at different depths (**Figure 20(e)–20(l)**) will complicate identifications of mantle upwellings in seismic images (Farnetani and Samuel, 2005; Kumagai et al., 2008; Lin and van Keken, 2006b,c), independent of how chemically dense materials (such as eclogite) influence seismic wave velocities (cf. Abalos et al., 2011; Connolly and Kerrick, 2002; Xu et al., 2008; Zhang and Green, 2007).

Further complexities arise from the interaction of thermochemical upwellings with mantle phase transitions. Farnetani and Samuel (2005) showed that the endothermic phase transition (and the associated viscosity jump) between the lower mantle and the upper mantle can alter thermochemical plume morphologies, for example, by promoting spout morphologies in the upper mantle (**Figure 20(a)–20(d)**). Furthermore, the density maximum of eclogite relative to peridotite in the depth range of 300–410 km (Aoki and Takahashi, 2004) can cause mantle upwellings to intermittently stall in the mid-upper mantle (Ballmer et al., 2013b; **Figure 20(m)**), which as discussed in **Section 7.10.2.5.2** can address the seismic structure of the Hawaiian plume (Cheng et al., in press; Wolfe et al., 2009).

The characteristics of thermochemical plume dynamics are relevant for our understanding of hotspot and flood basalt volcanism. The gravitational pull exerted by high-density eclogites in thermochemical plumes may account for the lack of

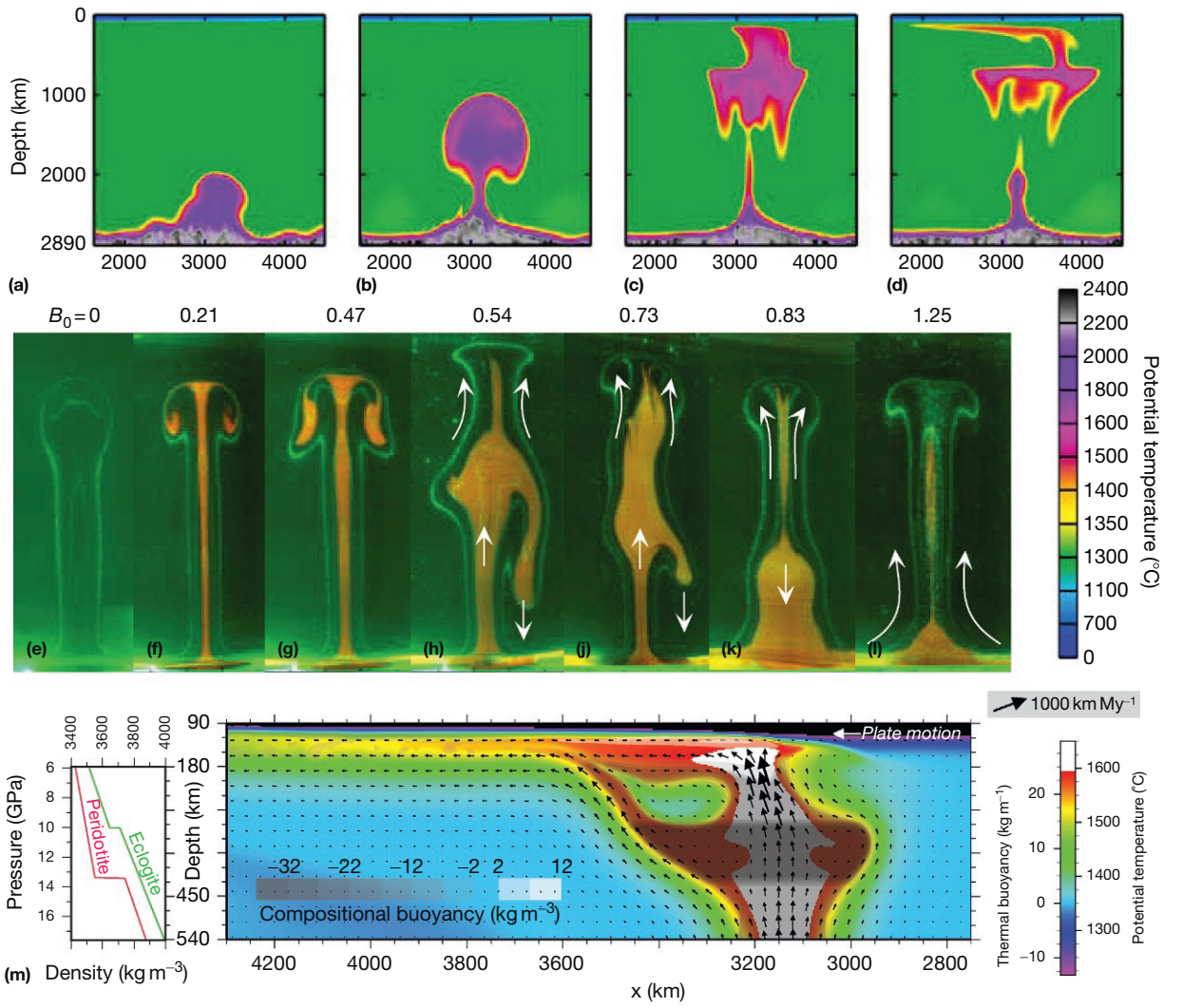


Figure 20 Complexity of thermochemical plume dynamics in vertical cross sections through three-dimensional geodynamic models. (a–d) Time series of a plume rising from a deep thermochemical boundary layer and interacting with a viscosity jump at 660 km depth. Colors denote temperature (scale displayed on central right). (e–l) Diverse plume shapes (as snapshots) for different compositional densities of the dense material (i.e., different buoyancy numbers B_0). The dense material is light orange; isotherms are light green. The case with $B_0 = 0$ (e) corresponds to a purely thermal upwelling. (m) Snapshot of a thermochemical plume in the upper mantle. The ascending plume with an eclogite content of 15% interacts with phase changes that modulate the excess density of eclogite relative to peridotite (see inset at bottom left as modified from Aoki and Takahashi (2004)). Potential temperatures and compositional buoyancy are displayed as colors and shadings, respectively. Black arrows show mantle flow velocities. (a–d) Reproduced from Farnetani CG and Samuel H (2005) Beyond the thermal plume hypothesis. *Geophysical Research Letters* 32: L07311, http://dx.doi.org/07310.01029/02005GL022360. (e-l) Reproduced from Kumagai I, Davaille A, Kurita K, and Stutzmann E (2008) Mantle plumes: Thin, fat, successful, or failing? Constraints to explain hotspot volcanism through time and space. *Geophysical Research Letters* 35. (m) Reproduced from Ballmer MD, Ito G, Wolfe CJ, and Solomon SC (2013b) Double layering of a thermochemical plume in the upper mantle beneath Hawaii. *Earth and Planetary Science Letters* 376: 155–164.

premagmatic uplift for the Siberian Traps (Sobolev et al., 2011) and other flood basalt provinces (see **Section 7.10.2.3**). Oscillatory plume behavior may not only explain the episodicity of LIP emplacement (Lin et al., 2005) but also reconcile magmatic-flux variations evident in the geologic record along many hotspot tracks (e.g., Adam et al., 2007; Mjelde et al., 2010; Zhang et al., 2011) such as Hawaii (van Ark and Lin, 2004; Vidal and Bonneville, 2004) and Iceland (Ito et al., 1999; Parnell-Turner et al., 2013; Poore et al., 2011). Finally, small plumes are predicted to rise intermittently from large, stagnant thermochemical domes (Davaille, 1999) to provide an explanation for short-lived hotspot activity as evident in the S and W Pacific (Clouard and Bonneville, 2005; Courtillot et al., 2003; Koppers et al., 2003).

7.10.3.3.4 Effects of variable mantle properties on plume dynamics

In addition to thermochemical effects, variations in mantle material properties can lead to complex forms and time dependence of mantle plumes. For example, a combination of increasing ambient mantle viscosity and thermal conductivity and decreasing thermal expansivity with depth has been shown

to cause plumes to be relatively broad in the deep mantle but become thinner when migrating upward (Albers and Christensen, 1996; Leng and Gurnis, 2012). Several effects that are often ignored in geodynamic models, such as those related to radiative heat transport (Matyska et al., 1994), the postperovskite transition (Matyska and Yuen, 2005, 2006), and/or the iron spin-state crossover (Matyska et al., 2011), also tend to increase plume widths in the deep mantle, perhaps to extents comparable to thermochemical domes.

The largest variations in mantle material properties occur at the upper mantle-to-lower mantle transition at 660 km depth. The sharp decrease in viscosity from the lower to upper mantle causes a sudden drop in the width of rising plumes, as well as a related increase in ascent rate (Kumagai et al., 2007; Leng and Gurnis, 2012; van Keken and Gable, 1995). The density contrasts associated with the 660 km discontinuity can further contribute to acceleration and episodicity of plume flow (Brunet and Yuen, 2000; Nakakuki et al., 1997). Tosi and Yuen (2011) showed that plumes with strongly temperature-dependent rheology may be deflected horizontally just below the 660 km discontinuity, which itself has been identified as a potential source region for plumes (Cserepes and Yuen, 2000).

An important aspect of mantle rheology is that under high-strain-rate conditions, the rheology is dominated by non-Newtonian effects with a nonlinear dependence of viscosity on stress. Such a rheology can dramatically enhance the deformation rate of boundary layer instabilities and lead to much higher rise speeds than is observed in Newtonian fluids (Larsen and Yuen, 1997; Larsen et al., 1999; van Keken, 1997). Starting plume heads can rise sufficiently fast to almost completely separate from the narrower (and thus slower) tail, a behavior that may form an alternative explanation for the observed LIP episodicity (van Keken, 1997). In turn, nonlinear yield-stress rheology has been shown to promote spoutlike instead of mushroomlike plume morphologies (Davaille et al., 2013; Massmeyer et al., 2013).

In summary, the classical descriptions of plumes as cylindrical, axisymmetric, continuously rising features from the CMB to the lithosphere captured many of the first-order physical processes and thus provided a powerful framework for understanding localized upwellings that give rise to hotspots. New observations and more sophisticated models, however, have shown that plume-like upwellings are likely to take on complex morphologies, be time-dependent, and originate from different depths in the mantle. Systematic comparison of model predictions with observations, as well as new constraints from mineral physics, is required to identify the range of behaviors that are most applicable to the Earth's mantle.

7.10.3.3.5 Small-scale sublithospheric convection

Beyond the complexities of active plume ascent, other forms of upwelling have been proposed to sustain intraplate volcanism. While plumes rise from a hot thermal boundary in the deep mantle, passive upwelling can be triggered by instabilities dripping down from the cold thermal boundary layer at the base of the lithosphere. Analogous to bottom-up Rayleigh–Bénard instabilities discussed earlier, such top-down-driven small-scale convective instabilities are promoted by the thickening of the cool material at the base of the lithosphere and low viscosities of the underlying asthenosphere. Small-scale convection (SSC) has been shown to be typically organized as rolls (**Figure 21(a)**), aligned by the direction of plate motion (Richter, 1973). Rolls are at first limited to the asthenosphere with typical wavelengths of 200–300 km but may extend deeper beneath mature plates (Korenaga and Jordan, 2004). In both cases, SSC acts to remove the cool base of the lithosphere and to replace it by warm materials from below.

Such a process can provide an explanation for the observed flattening of seafloor bathymetry (and of surface heat flux) in ocean basins of ages >70 Ma (Davaille and Jaupart, 1994; Doin and Fleitout, 1996; Stein and Stein, 1994; Zlotnik et al., 2008). SSC beneath young seafloor, however, may be restricted due to the presence of the residual harzburgite from previous MOR melting (Afonso et al., 2008; Hirth and Kohlstedt, 1996; Lee et al., 2005). The occurrence of SSC can further be tested using seismic observables (cf. Sleep, 2011). For example, it is consistent with the seismically derived thermal structure of the Pacific lithosphere (Ritzwoller et al., 2004; van Hunen et al., 2005).

The removal of the base of the cool (and depleted) lithosphere by SSC may moreover act to induce decompression melting (**Figure 21(b)**; Ballmer et al., 2007; Buck and Parmentier, 1986). According to the geometry typical for SSC (**Figure 21(a)**), magmatism is predicted to emerge along one or multiple hot lines parallel to the direction parallel to plate motion (Bonatti and Harrison, 1976; Bonatti et al., 1977), rather than at localized hotspots. Associated volcanic activity can thus occur coevally over distances of up to 1000–2000 km (Ballmer et al., 2009). Such extensive volcanism can create complex age–distance patterns such as those observed among various seamount chains in the western and southern Pacific (cf. shaded fields in **Figure 5**).

An important limitation for this mechanism is that significant decompression melting requires SSC to occur beneath thin lithosphere of moderate or young ages (Ballmer et al., 2007; Buck and Parmentier, 1986; Harmon et al., 2011; Haxby and Weissel, 1986). Seafloor flattening, instead, indicates that SSC is typically restricted to thick lithosphere of ages >70 Ma, mostly due to the tendency for the dehydrated harzburgitic residue of MOR melting to delay SSC (Afonso et al., 2008; Hirth and Kohlstedt, 1996; Lee et al., 2005). Accordingly, SSC would likely have to stir up fertile or excessively hot materials to feed significant volcanism (Ballmer et al., 2010). Alternatively, SSC may be advanced toward younger lithospheric ages near fracture zones (Dumoulin et al., 2008; Huang et al., 2003).

Adjacent to larger steps in lithospheric thickness, convection is expected to be triggered almost independently of the thickness of the thermal boundary layer. This form of 'edge-driven' thermal convection can spawn decompression melting along rifts, as well as parallel to the edges of thick continental lithosphere (**Figure 22**) with opportunities for volcanism (Buck, 1986; Hardebol et al., 2012; King, 2007; King and Anderson, 1998; King and Ritsema, 2000; Missenard and Cadoux, 2012; Till et al., 2010; van Wijk et al., 2010). Recently, Milelli et al. (2012) showed in analogue experiments that convective instability may also form perpendicular to continental margins, a configuration that could provide an explanation for mantle melting beneath the Cameroon volcanic line (cf. **Figure 6**). Edge-driven convection along rifted margins, alone or in combination with elevated mantle temperatures,

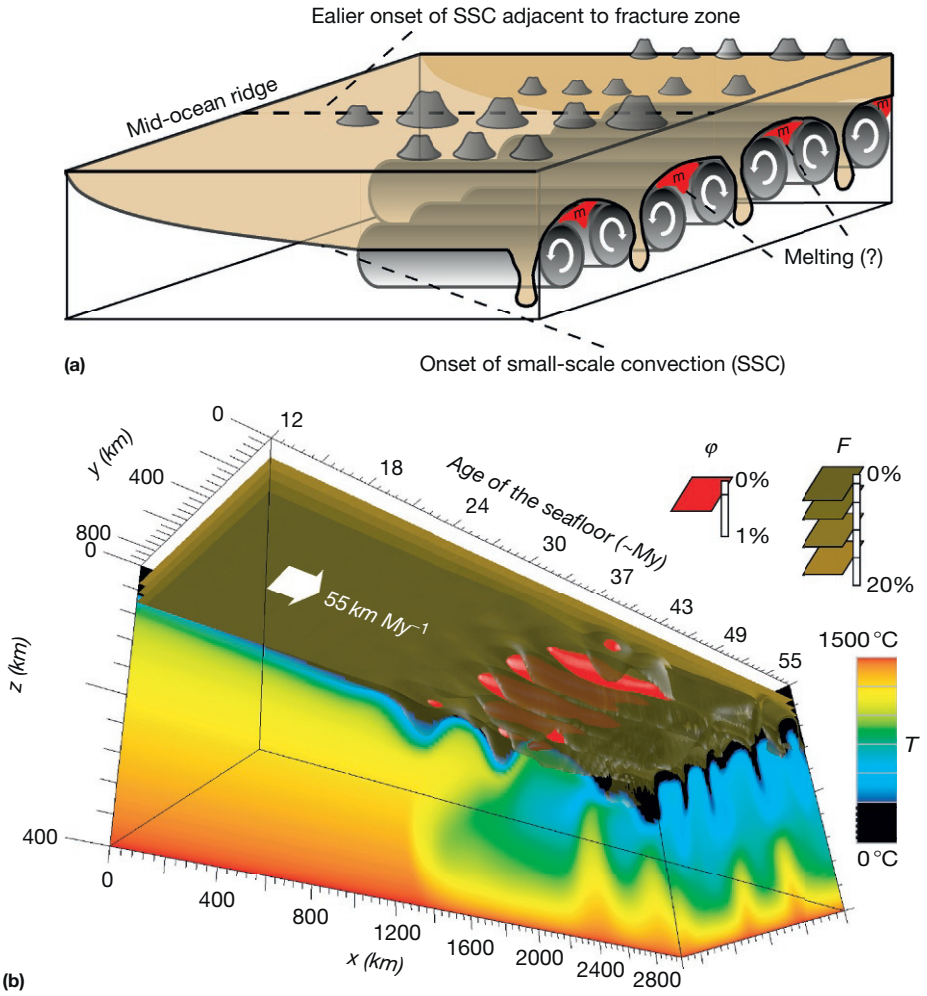


Figure 21 (a) Conceptual cartoon of the dynamics of small-scale convection (SSC). Any decompression melting ('*m*') associated with SSC is expected to occur along multiple parallel 'hot lines' associated with the upwelling limbs of SSC. (b) Temperature, melting, and depletion predicted by a three-dimensional geodynamic model of SSC. Retained melt (φ) is labeled by red, and depletion in peridotite (*F*) by brown isosurfaces. Rainbow colors show temperature (*T*) along two cross sections. The flow field is not shown but is similar to that sketched in (a). SSC dynamically removes the base of the cool harzburgitic layer (brown) and replaces it by warm and undepleted peridotite to spawn decompression melting (red). Reproduced with minor modifications from Ballmer MD, van Hunen J, Ito G, Bianco TA, and Tackley PJ (2009) Intraplate volcanism with complex age-distance patterns: A case for small-scale sublithospheric convection. *Geochemistry, Geophysics, Geosystems* 10.

has even been proposed to be a contributing factor in flood basalt volcanism (Korenaga, 2004; Mutter and Zehnder, 1988; Nielsen and Hopper, 2004; Nielsen et al., 2002; Sleep, 2007). Spatially and temporally more restricted convective processes such as the return flow associated with local foundering and delamination of the lower continental lithosphere have also been proposed to account for intraplate volcanism (e.g., in western North America) and flood volcanism without elevated mantle temperatures (Crow et al., 2011; Elkins-Tanton, 2007; Gogus and Pysklywec, 2008a,b; Hales et al., 2005; Houseman et al., 1981; Le Pourhiet et al., 2006; Levander et al., 2011; Paczkowski et al., 2012; Reid et al., 2012; van Wijk et al., 2001; West et al., 2009).

7.10.3.3.6 Buoyant decompression melting

A separate form of convective instability can occur in response to melting itself. Partial melting reduces the density of the solid residue (Jordan, 1979; Oxburgh and Parmentier, 1977; Schutt and Lesher, 2006) and generates intergranular melt. Both factors can reduce the bulk density of the partially molten mantle and drive 'buoyant decompression melting' (**Figure 23**). Buoyant decompression melting has been shown to be relevant near MORs, for example, in supporting off-axis melting anomalies (e.g., Jha et al., 1994; Scott and Stevenson, 1989; Sparks et al., 1993; VanDecar et al., 1995). It has also been proposed to be relevant for intraplate volcanism, on both continental and oceanic plates (Hernlund et al., 2008a,b; Raddick et al., 2002; Tackley and Stevenson, 1993). The key ingredients for buoyant decompression melting to occur well away from MORs are an asthenosphere close to or at its solidus and a perturbation that locally initiates melting. Such a perturbation could be caused by extension of the overlying lithosphere (Hernlund et al., 2008b), the flow of mantle across a step in lithospheric thickness (Raddick et al., 2002), or some form of passive mantle

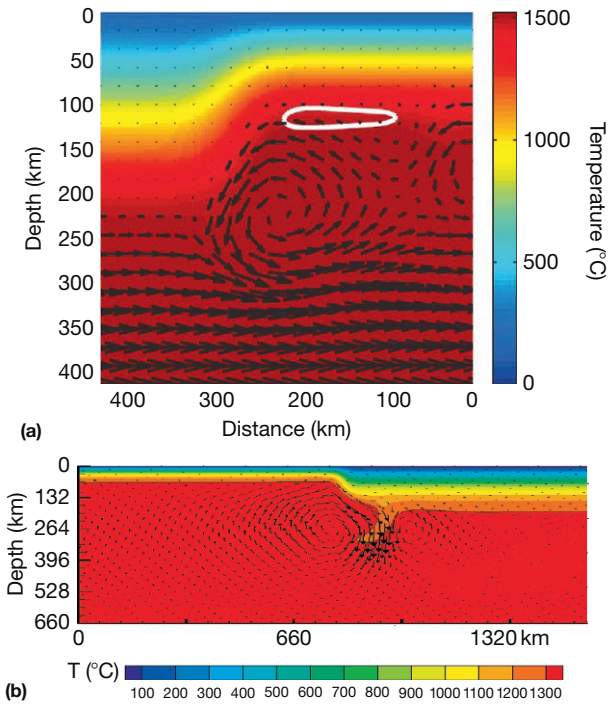


Figure 22 Numerical model predictions for two different simulations of edge-driven convection along steps in lithospheric thickness. Colors represent temperatures, and arrows mantle flow velocity vectors. Mantle flow in (a) has a strong component of horizontal shear flow that contributes to drive upwelling (cf. **Figure 24(a)**). White line highlights the mantle domain, in which the solidus temperature is exceeded and melting may occur. Upwelling in (b) is predominantly passively driven by delamination near the sublithospheric step. In both (a) and (b), upwelling occurs >100 km away from the step in lithospheric thickness. (a) Reproduced from Till CB, Elkins-Tanton LT, and Fischer KM (2010) A mechanism for low-extent melts at the lithosphere-asthenosphere boundary. *Geochemistry, Geophysics, Geosystems* 11. (b) Reproduced from van Wijk JW, Baldridge WS, van Hunen J, et al. (2010) Small-scale convection at the edge of the Colorado Plateau: Implications for topography, magmatism, and evolution of Proterozoic lithosphere. *Geology* 38: 611–614.

flow. Significant retention of buoyant melts and low mantle viscosities tend to promote this convective instability. Since higher mantle water contents act not only to reduce the viscosity of the solid mantle but also to mobilize the equilibrium liquid, its effects on buoyant decompression melting remain unclear and thus motivate future studies.

7.10.3.3.7 Shear-driven upwelling and viscous fingering

All processes for decompression melting described so far require an inversion in density stratification (e.g., cool mantle overlying hot mantle) as the (potential) energy source of upwelling. Alternatively, vertical flow can result from horizontal shearing of the asthenosphere independent of mantle density variations (Bianco et al., 2011a; Conrad et al., 2010). Asthenospheric shearing is commonly imposed by the motion of the tectonic plates relative to that of the mesosphere and can drive upwelling in two different situations. First, (1) horizontal shear flow is predicted to be guided upward (or downward) at a step in lithospheric thickness (**Figure 24(a)**). Second, (2) redistribution of horizontal shear flow due to strong viscosity heterogeneity is predicted to be accommodated by local upwellings and downwellings (**Figure 24(b)**); vertical motion arises as shear flow is guided around a high-viscosity anomaly or focused within a low-viscosity anomaly.

Shear-driven upwelling due to the presence of a low-viscosity anomaly (situation (2); **Figure 24(b)**) is an attractive scenario for intraplate volcanism, since such anomalies are likely to be damp or warm and therefore close to the solidus. Regions of anomalously low seismic-wave speeds in the upper mantle (Laske et al., 2011; Schmandt and Humphreys, 2010; Weeraratne et al., 2007) provide evidence for low-viscosity pockets due to excess temperature, partial melt, elevated water content, or a combination thereof. Shear-driven upwelling at the edges of such pockets can support moderate magmatism (**Figure 24(c)**). Associated age progressions are predicted to be similar to those of (rapidly) drifting hotspots but with more variability (Ballmer et al., 2013a; Bianco et al., 2011a), similar to that evident along, for example, the Pukapuka ridge (cf. **Figure 4(c)**). Global mantle convection models predict

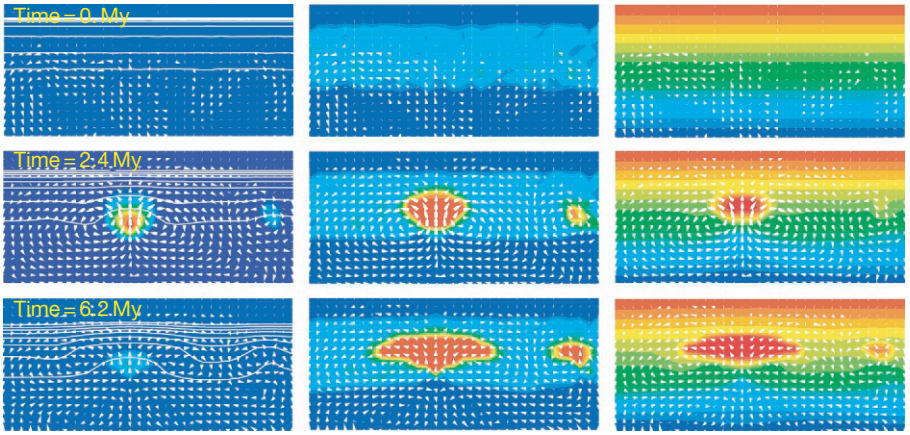


Figure 23 From Raddick et al. (2002). Predictions of 2-D numerical models that simulate buoyant decompression melting for 3 time steps (different rows as labeled). Left column shows fractional melting rate (red = 0.0021/My, blue = 0), middle column shows melt fraction retained in the mantle φ (red = 0.02), and right column shows depletion of the residue F (red = 0.108). Density decreases as linear functions of both φ and F, and lateral density variations are what drive upwelling. Melting is therefore limited by the accumulated layer low-density and high-solidus residue.

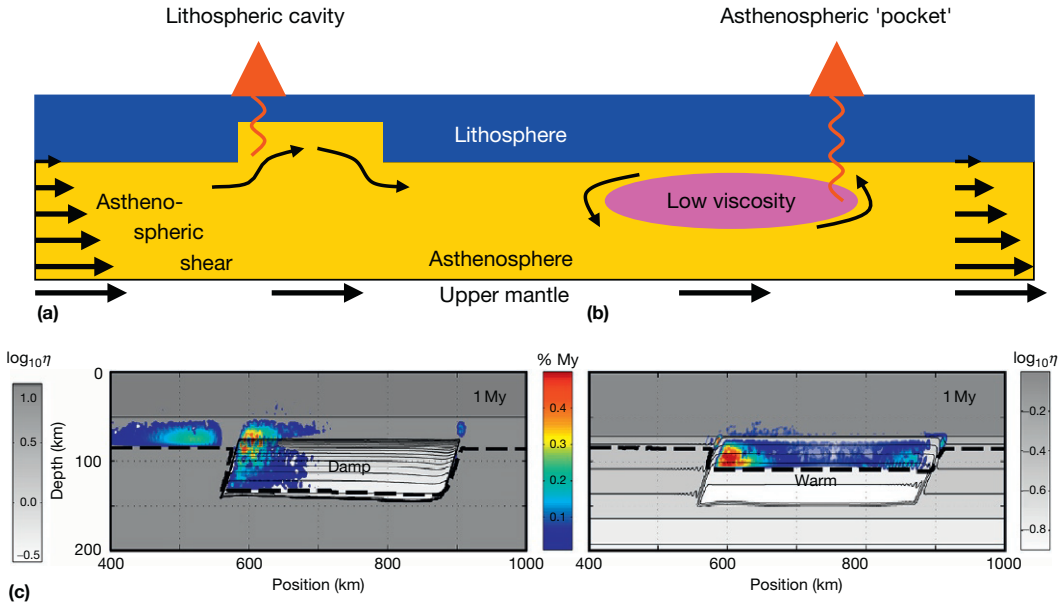


Figure 24 Adapted from (a–b) http://www.mantleplumes.org/ShearDrivenUpwelling.html; and reproduced from (c) Bianco TA, Conrad CP, and Smith EI (2011a) Time dependence of intraplate volcanism caused by shear-driven upwelling of low-viscosity regions within the asthenosphere. *Journal of Geophysical Research* 116. (a–b) Cartoon showing two end-member scenarios, in which shear-driven upwelling is expected to occur (see text). Horizontal shear flow in the asthenosphere may be redirected into vertical flow when entering a cavity (a) or a pocket of anomalously low viscosity (b). (c) Snapshot of two numerical models (model time 1 My) quantifying the effects of shear-driven upwelling for the scenario shown in (b). Both damp and warm pockets locally reduce the viscosity η (shading and contours) to induce upwelling (not shown) and decompression melting (colors).

that intraplate regions with excessive intraplate volcanic activity, such as the eastern Pacific (**Figure 4**), western North America, and northeastern Australia (**Figure 7**), are indeed underlain by a strongly sheared asthenosphere (Conrad et al., 2011).

Identifying shear-driven upwelling as the dominant mechanism for decompression melting in a given volcanic province, however, requires careful analysis of the tectonic setting. In both situations (1) and (2), in which shear-driven upwelling is predicted to occur (near steps in lithospheric thickness and low-viscosity anomalies; **Figure 24(a)** and **24(b)**), alternative mechanisms for decompression melting, such as buoyant decompression melting and (edge-driven) small-scale convection, are also expected to be relevant (e.g., Huang et al., 2003). Increasing asthenospheric background viscosities and/or the rate of asthenospheric shearing promotes shear-driven upwelling relative to convective instabilities.

Future efforts should be specifically directed at testing the general significance of shear-driven upwelling along the margins of low-viscosity anomalies (situation (2); **Figure 24(b)**). Geophysical studies are needed to assess whether anomalies imaged by tomography can provide sufficient viscosity contrasts (i.e., 1–2 orders of magnitude at a given rate of asthenospheric shearing). Whereas deformation of the pocket in the large-scale flow field has been shown to be less of an issue (cf. Ballmer et al., 2013a; Bianco et al., 2011a), the persistence of magmatism should be limited by the progressive consumption of volatiles or fertile rocks in the pocket. Accordingly, it is imperative to address the origin of the pocket together with that of melting.

A mechanism that can spawn viscosity heterogeneity in the asthenosphere and that may even lead to excess volcanism by itself has been suggested to be operant in the South Pacific. When two fluids are contained in a thin layer and one fluid displaces a more viscous fluid (driven by a pressure gradient), the boundary between the two becomes unstable and undulates with increasing amplitude (Saffman and Taylor, 1958). Such viscous fingering instabilities may occur in nature, where warm and fertile material diffusely rising beneath the South Pacific Superswell is injected into the ambient asthenosphere. This process may account for the apparent 'Haxby' gravity lineations (Goodwillie, 1995; Harmon et al., 2011; Haxby and Weissel, 1986; Sandwell et al., 1995; Weeraratne et al., 2007) and associated volcanism (e.g., along the Pukapuka ridge) as warm and buoyant fingers travel along the base of the oceanic lithosphere (Sleep, 2008; Weeraratne et al., 2007; **Figure 4**). Whereas related rates of decompression melting should be rather small beneath the quasiflat lithospheres of fast or old plates, a combination with shear-driven upwelling at the tips of the fingers is expected, as both mechanisms are driven by similar physics.

Similarly, material injected into the asthenosphere by mantle plumes (Phipps Morgan et al., 1995b; Yamamoto et al., 2007) may feed warm low-viscosity channels. Geodynamic models predict that the pressure low expected beneath MORs can assist in drawing material from nearby plumes toward ridges in relatively narrow channels (e.g., Sleep, 1996; Yale and Phipps Morgan, 1998), as was originally hypothesized by Morgan (1978). Channeling of plume material is consistent with hotspot-flavored geochemical fingerprints of basalts formed near or at MORs (Geldmacher et al., 2013; Haase et al., 1996; Niu et al., 1999; Vlastelic and Dosso, 2005), as well as between the Canary hotspot and a subduction zone near Gibraltar (see Duggen et al. (2009) and **Figure 25**), respectively.

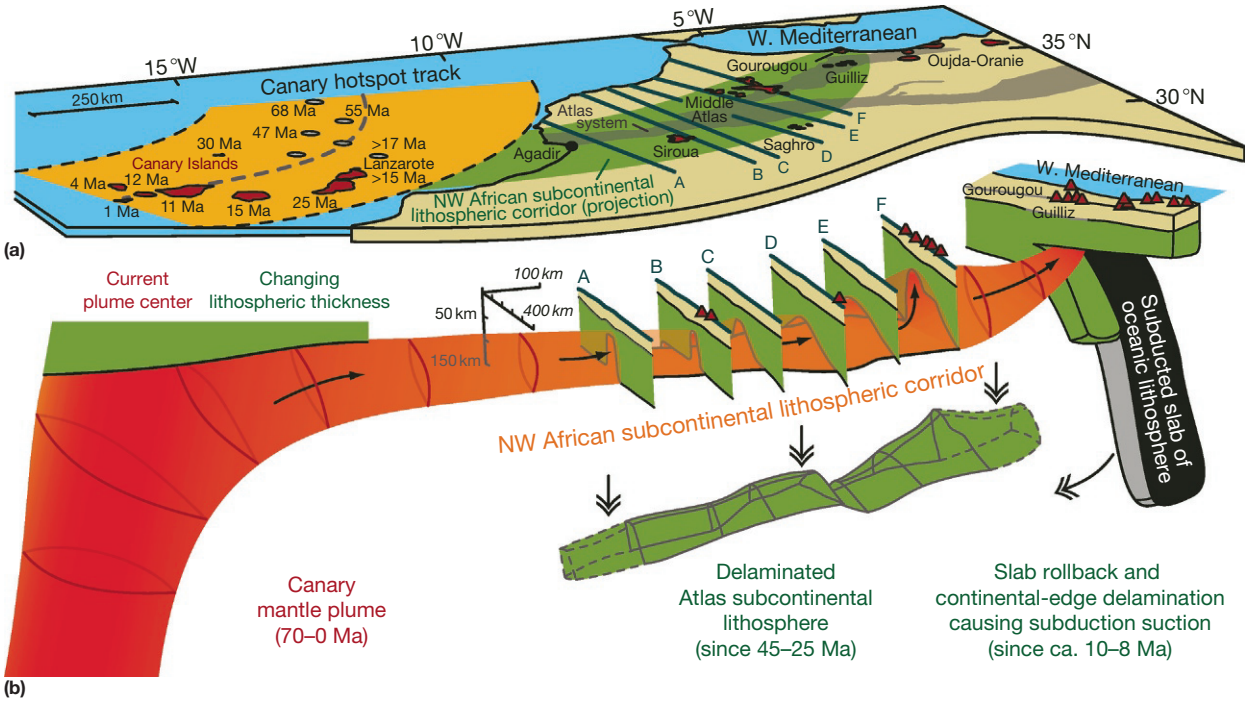


Figure 25 From Duggen et al. (2009). Conceptual model for the channeling of plume material from the Canary hotspot through the northwest African subcontinental lithospheric corridor to the mantle wedge of a subduction zone near Gibraltar. Upward flow along the base of the lithosphere, perhaps assisted by shear-driven upwelling (cf. **Figure 24**), may have contributed to sustain Cenozoic basaltic volcanism along the corridor (red triangles in (b)).

7.10.3.3.8 Vertical motion of the lithosphere

Finally, vertical motion of the whole lithosphere and surface of the Earth can cause mantle upwelling and melting. For example, volcanism off southern Greenland (Uenzelmann-Neben et al., 2012) and peaks in volcanic activity in Iceland (Slater et al., 1998) have been attributed to glacial rebound. Another opportunity for mantle decompression is offered by the growth of large volcanoes (created by another mechanism such as a plume), which not only causes the underlying lithosphere to sink downward but also leads to flexure-induced upwarping in a donut- or arch-shaped zone around the volcano. The small amounts of decompression associated with this flexural uplift are only expected to give rise to magmatism as long as an appreciable portion of the asthenosphere is already near or at its solidus (Bianco et al., 2005). Flexural decompression may thus explain secondary (or rejuvenated-stage) volcanism near the Hawaiian and other major hotspots (e.g., Bianco et al., 2005; Garcia et al., 2010). Upward flexural bulging also occurs on the seaward side of subduction zones and may sustain the activity of 'petit-spot' volcanism off the Japan Trench (Hirano et al., 2001, 2008), as well as contribute to extensive rejuvenated volcanism at the Samoan hotspot (Konter and Jackson, 2012).

7.10.3.4 Swells

7.10.3.4.1 Swell support mechanisms

The origin of hotspot swells was initially attributed to heat anomalies in the mantle that effectively rejuvenate or thin the overriding lithosphere (Crough, 1978; Detrick and Crough, 1978; Crough, 1983). The evidence for such thermal rejuvenation is somewhat ambiguous, as heat flow data fail to show evidence for heated or thinned lithosphere (DeLaughter et al., 2005; Stein and Stein, 1993, 2003). However, it is likely that any such evidence would be obscured by shallow hydrothermal circulation (Harris and McNutt, 2007; McNutt, 2002). Seismic studies, in turn, are able to provide evidence for lithospheric thickness variations at some hotspots. For example, Rayleigh-wave tomography reveals strong variations in lithospheric thickness at the Galápagos hotspot (Villagomez et al., 2007). At the Hawaiian hotspot, S-wave receiver functions indicate substantial lithospheric thinning of 50–60 km from the hotspot toward the NW (Li et al., 2004), whereas recent evidence from regional surface-wave tomography as well as SS precursors suggests only modest thinning (by about 10–20 km) (Laske et al., 2011; Schmerr, 2012).

Perhaps, the more prevalent view at this point is that hotspot swells are primarily dynamically supported by the hot and buoyant plume layer that is predicted to pond beneath the lithospheric plate (Cadio et al., 2012; Cserepes et al., 2000; Olson, 1990; Parsons and Daly, 1983; Ribe and Christensen, 1994). For example, dynamic support by anomalously hot mantle (as evident from seismic tomography) can explain excess topography associated with the Yellowstone hotspot (Becker et al., 2013). In addition to the thermal buoyancy of a plume, compositional buoyancy associated with melt extraction may also be important in creating hotspot swells (Phipps Morgan et al., 1995a).

Evidence for such compositional contributions to swell support has been revealed at various hotspots. Exposures of harzburgite along the rifted Marion and Azores Rises, for

instance, indicate compensation by buoyant melt residuum (Zhou and Dick, 2013). Uplift histories and receiver function measurements argue for a combination of buoyant melt residuum and magmatic underplating to support the ~2 km high Cape Verde swell (Lodge and Helffrich, 2006; Ramalho et al., 2010). For the Marquesas swell, geoid observations and seismic reflections also suggest a role for magmatic underplating (Caress et al., 1995; McNutt and Bonneville, 2000).

A viable tool to quantify the compensation depth of the swell and thus to distinguish between crustal (e.g., magmatic underplating), lithospheric (e.g., lithospheric thinning), and mantle (e.g., dynamic support and buoyant melt residuum) contributions is the analysis of the geoid-to-topography ratio (Marks and Sandwell, 1991; Parsons and Daly, 1983; Ribe and Christensen, 1994). For example, once the local effects of volcanic loading are removed (cf. Cserepes et al., 2000), wavelet analysis of the geoid-to-topography ratio reveals that the Hawaiian swell is predominantly compensated at sublithospheric depths (Cadio et al., 2012). The best candidate for such a deep compensation is dynamic support by a mantle plume.

7.10.3.4.2 Plume models and hotspot swells

The dynamic support hypothesis for the generation of hotspot swells has been quantitatively tested by geodynamic models. Three-dimensional plume models, for example, have successfully predicted the shape and uplift history of the Hawaiian swell (Asaadi et al., 2011; Cadio et al., 2012; Ribe and Christensen, 1994; van Hunen and Zhong, 2003; Zhong and Watts, 2002): they predict that the plume pushes up the plate by ponding like a 'pancake' of hot material (cf. red layer in **Figure 26(a)**). They also predict the eventual waning of swell topography to occur due to spreading and thinning as well as cooling of the plume pancake. Consideration of stress-dependent rheology appears to be critical to match the plan-view shape of the swell to the NW of Hawaii (Asaadi et al., 2011). The additional effects of small-scale convection in the pancake removing the base (about 15–20 km) of the lithosphere (Agrusta et al., 2013; Ballmer et al., 2011; Moore et al., 1998) can account for the systematic decrease of the geoid-to-topography ratio that is evident along the Hawaiian swell (Cadio et al., 2012). These predictions provide a straightforward explanation for the decay of hotspot swells along the Hawaiian and Louisville chains as well as the lack of swells around very old portions of other volcano chains.

7.10.3.5 Large Igneous Provinces

The rapid and massive magmatic production of LIPs, combined with their strong connection to continental breakup and present-day hotspot volcanism, provides major challenges for our understanding of planetary volcanism. The wide range of eruptive volumes (**Figure 10**) and differences in tectonic style (**Section 7.10.2.3**) appear to suggest that multiple mechanisms account for flood basalt volcanism.

The observation of large plume heads followed by thin tails in fluid dynamic experiments has traditionally been used to explain LIP–hotspot connections (Richards et al., 1989) and remains, because of its simplicity and plausibility, an attractive base model for the formation of LIPs. Its strengths include that

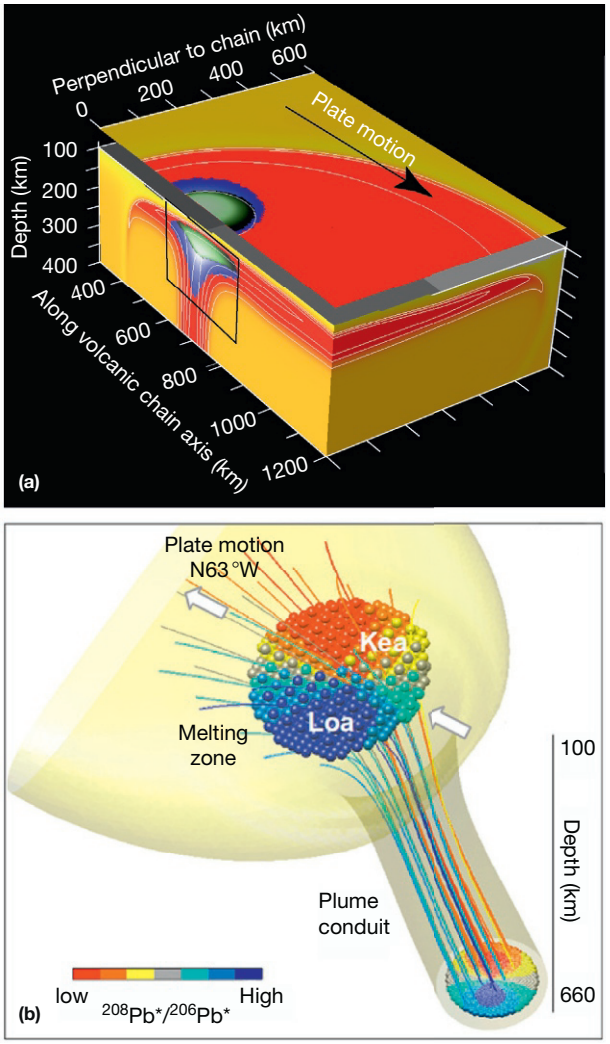


Figure 26 Three-dimensional images (in angled views from above) of plume dynamics and the expressions of compositional heterogeneity in magmatism. Whereas melting is affected by both temperature and composition in the geodynamic models shown, plume dynamics are governed by thermal effects only. (a) From Bianco et al. (2008). White contours (isotherms), as well as red-to-yellow colors, denote temperature. The blue and light green fields show the melting zones of a fertile and refractory compositional component, respectively. The compositional heterogeneity is assumed to be in the form of evenly dispersed fine-scale fertile veins in a refractory matrix. A horizontal cross section through the box at 150 km depth is lifted from the top for visibility. (b) From Farnetani et al. (2012). The isosurface outlines the hot plume; the colored lines denote streamlines tracking large-scale compositional heterogeneity. Heterogeneity is assumed to be distributed bilaterally asymmetrically in the deep plume stem (as represented by $^{208}Pb^*/^{206}Pb^*$ signatures in a cross section at 660 km depth (colors)). The expression of heterogeneity in magmatism is expressed by sphere colors.

(1) it is supported by fluid dynamic experiments with increasingly realistic assumptions about mantle composition and rheology (see **Section 7.10.3.3**; in fact, modifications to the base model (cf. **Figure 20**) can reconcile the diversity seen in the geologic record), (2) it can explain the natural link between LIP and hotspot volcanism, (3) it reconciles the early occurrence of high-MgO basalts in the LIP record (as the hottest

material is predicted to erupt first), and (4) it implies contemporaneous uplift and extension as is observed in the geologic record of many LIPs (see **Section 7.10.2.3**). The main weaknesses of this model involve that the strong correlation of LIPs and continental breakup and the lack of uplift during OJP formation (d'Acremont et al., 2003) are not explained.

There are a number of alternatives that can address some of the weaknesses of the plume model that include assumptions that LIPs may be generated by processes occurring near the surface of the Earth or by extraterrestrial processes. While the plume model has received significant attention and quantitative hypothesis testing, the majority of the alternatives are currently still in rather qualitative form.

The first alternative is that the accumulation of heat below the insulating continent or supercontinent during tectonic quiescence provides warmer than normal mantle and significantly larger eruptive volumes during continent breakup (Anderson, 1994b; Coltice et al., 2007; Grigne et al., 2007; Coltice et al., 2009). While this hypothesis addresses the correlation between LIPs and continental breakup, it does not clarify why flood volcanism does not generally occur along all rifted margins. Also, the connection to a long-lived hotspot trail remains unpredicted. Nevertheless, the correlation between the LIPs and the continental breakup is intriguing, and it is quite likely that regional variations in the composition and strength of the lithosphere have an important control on the location of magma eruption (King and Anderson, 1995, 1998).

A second alternative is that compositional, rather than thermal, effects cause the generation of excessive melt by, for example, the presence of eclogite or volatiles in the source rock (e.g., Anderson, 1994a, 2005; Cordery et al., 1997; Korenaga, 2004, 2005; Sobolev et al., 2011). Its strengths include that the lack of uplift at some LIPs, including the OJP, could be explained for the presence of dense eclogite. The main weakness of eclogite as the primary source of melting is that eclogite is dense and requires some mechanism to stay near or be brought back to the surface. It would be more than likely in this case that emplacement of the OJP involved both a thermal origin and a compositional origin. For purely compositional effects, the common LIP–hotspot connection (requiring continuous upwelling of more fertile mantle) further remains unexplained. An interesting alternative to a compositional cause for the lack of significant uplift at some LIPs, including the OJP, is the potential for lower crustal flow during magma emplacement (Flament et al., 2011).

The enigmatic nature of the OJP has also led to the suggestion that meteorite impacts could have been responsible for LIP emplacement (Ingle and Coffin, 2004; Rogers, 1982). A possible strength of this hypothesis is that the decompression of mantle following impact may generate extensive melting (Jones et al., 2002a). A weakness is that the strong connections with continental breakup and hotspot volcanism are unexpected. Furthermore, direct evidence for meteorite impact during LIP emplacement is rare. One of the few convincing observations is the Ir anomaly embedded in the Deccan Traps, but this impact signal postdates the start of volcanism and is most likely related to the Chicxulub impact located on the other side of the planet. There are also doubts whether the dynamics of meteorite impact can indeed cause sufficient melting, unless the mantle is already anomalously hot or fusible (Ivanov and Melosh, 2003). Another issue is that it has been shown to be statistically unlikely that the majority of Phanerozoic LIPs originated as impacts (Elkins-Tanton and Hager, 2005; Ivanov and Melosh, 2003).

Finally, delamination of continental lithosphere and secondary convection at rifted margins have been suggested as alternatives for LIP formation (Anderson, 2005; Hales et al., 2005; van Wijk et al., 2001). While these processes are likely to have occurred regularly in the past, the extent to which they can provide a consistent explanation for the rapid and massive outpourings of flood basalts remains to be quantitatively evaluated.

7.10.3.6 The Origin of OIB Geochemical Signatures

7.10.3.6.1 Tracing mantle heterogeneity by isotopic signatures of OIB

The geochemical signatures of ocean island basalts (OIBs) may provide key constraints on the composition of mantle reservoirs. Highlighting the differences between OIB and MORB, we have focused on $^{87}Sr/^{86}Sr$, $^{206}Pb/^{204}Pb$, and $^{3}He/^{4}He$ (**Table 1**). These ratios are key to tracing five different geochemical flavors in the OIB dataset, which have been associated with five different mantle reservoirs, each with distinct time-averaged, chemical histories (e.g., Hanan and Graham, 1996; Hart et al., 1992; Zindler and Hart, 1986). Lavas with low $^{3}He/^{4}He$ and minimal $^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$ are derived from depleted mantle material (DM), referring to depletion in incompatible elements. Enriched (EM1 or EM2, i.e., high $^{87}Sr/^{86}Sr$) and HIMU (i.e., high $^{206}Pb/^{204}Pb$) mantle sources are instead thought to be derived from subducted material – the former from ancient oceanic sediments or metasomatized lithosphere and the latter from oceanic crust – that has evolved in the mantle for >1 Gyr and subsequently been recycled (e.g., Cohen and O'Nions, 1982; Hanyu et al., 2011; Hart et al., 1992; Hofmann, 1997; Hofmann and White, 1982; Sobolev et al., 2008; Zindler and Hart, 1986).

High $^{3}He/^{4}He$, moderately low $^{87}Sr/^{86}Sr$, and intermediate-to-high $^{206}Pb/^{204}Pb$ compositions mark the fifth geochemical end-member, which we will here refer to as FOZO (Hart et al., 1992). Its origin is perhaps least well understood. The 'standard' hypothesis implies that the high $^{3}He/^{4}He$ fingerprints primordial and relatively undegassed mantle material (Hart et al., 1992; see also Gonnermann and Mukhopadhyay, 2009). In a variant of this standard hypothesis, FOZO is not just a leftover of quasihomogenous primordial mantle, but has been enriched through a differentiation process occurring early in the Earth's history (Coltice et al., 2011; Davies, 2010; Lee et al., 2010). The standard hypothesis, however, has recently altogether been challenged by suggestions that FOZO, in fact, has been depleted in highly incompatible elements during partial melting processes. In this scenario, helium degasses only incompletely during partial melting (Class et al., 2005), and FOZO's high $^{3}He/^{4}He$ ratio reflects a low ^{4}He concentration as a result of low U and Th content (with U and Th being more incompatible than helium) (Coltice and Ricard, 1999; Meibom et al., 2005; Parman et al., 2005; Stuart et al., 2003). However, any alternatives to the standard hypothesis do not satisfy constraints from other noble-gas systems such as $^{129}Xe/^{130}Xe$ and $^{20}Ne/^{22}Ne$, which require that the relatively undegassed reservoir (i.e., FOZO)

formed within the first 100 million years of the Earth's history and since then has not been homogenized with the other mantle reservoirs (Mukhopadhyay, 2012; Peto et al., 2013).

The key observation is that MORB appears to be characteristically similar to DM and to be relatively less influenced by subducted materials and FOZO, whereas hotspot lavas and many other intraplate basalts appear to be influenced substantially by all five components (albeit to different degrees for different volcano groups; cf. **Figure 11**). The dominant explanation for this observation is that the pressure/temperature dependence of mantle viscosity and mineralogy, as well as density differences between the different mantle materials, promotes large-scale (partial) layering in mantle geochemistry. DM is likely to be compositionally light and may tend to concentrate in the upper mantle where it is sampled by MOR magmatism (cf. Nakagawa et al., 2010) Mantle plumes, which feed hotspots, rise from deeper levels in the mantle and hence may incorporate the other materials in addition to DM. For example, geochemical reservoirs associated with subducted slabs that evolved in the mantle for >1 Gyr (EM, HIMU) or even the more ancient FOZO are likely to be stored in the lowermost mantle (e.g., Brandenburg et al., 2008; Christensen and Hofmann, 1994; Coltice and Ricard, 1999).

7.10.3.6.2 Constraining source materials from major-element characteristics of OIB

While the various mantle reservoirs are well traced by their isotopic signatures in OIB, understanding their major-element signatures is needed to define their physical properties as well as melting behaviors. As a first step toward this goal, Jackson and Dasgupta (2008) demonstrated that major-element concentrations of OIB are correlated with isotopic ratios using a global dataset. At a given MgO, HIMU-flavored volcano chains are systematically lower in SiO_2 and Al_2O_3 and higher in TiO_2, Na_2O, FeO_X, and CaO than EM-flavored volcano chains, and both are distinct from MORB (**Figure 27**). Tholeiitic basalts from Hawaii, the Galápagos, and Iceland display major-element compositions intermediate to these two OIB subgroups and MORB. Thus, mantle reservoirs can in principle be distinguished not only by OIB isotopic signatures but also by OIB major-element concentrations.

To understand the composition of mantle reservoirs and OIB source lithologies, petrologists have compared experimental melts with typical alkalic OIB signatures. Although low-degree melting of 'common' mantle peridotites is able to produce alkalic melts, experimental evidence is so far lacking that such melts can match the specific major-element signature of OIB (Davis et al., 2011; Hirose and Kushiro, 1993; Mueller et al., 1998; Takahashi and Kushiro, 1983) Addition of CO_2-rich liquid to peridotites would be required to better match these signatures (blue triangles in **Figure 12**; Dasgupta et al., 2007). The characteristic HIMU isotopic fingerprint of most alkalic OIB further calls for a subduction-related origin for the CO_2-rich liquid (Jackson and Dasgupta, 2008).

Alternative materials for the OIB source are subduction-related lithologies such as basaltic oceanic crust that has transformed to eclogite and evolved in the mantle. While direct eclogite melting produces liquids that are much too high in SiO_2, two-stage eclogite melting can reconcile the major-element characteristics of OIB. In the first step, highly silicic

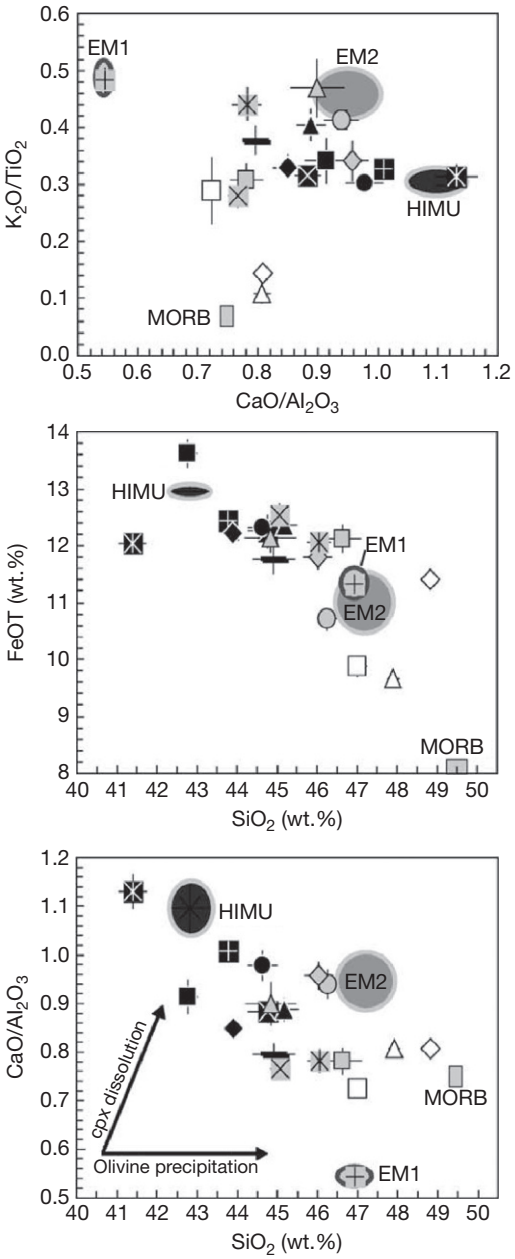


Figure 27 Separation of isotopically flavored OIB subsets in major-element space. Symbols (with ±1σ error bars) show select average major-element concentrations and ratios of fractionation-corrected lavas ($10\% \leq gO \leq 16\%$) from various island groups. EM-flavored and HIMU-flavored island groups are the gray and black symbols, respectively; their overall averages are represented by ellipses (Pitcairn–Gambier is the sole EM1-flavored island group). Open symbols represent the dominantly tholeiitic lavas from Hawaii, the Galápagos, and Iceland, which plot closer to MORB (cf. gray rectangle) than alkalic OIB. Arrows in (c) elucidate that neither dissolution of clinopyroxene nor precipitation of olivine can account for the global trend in CaO/Al_2O_3 vs. SiO_2. Reproduced from Jackson, and Dasgupta R (2008) Compositions of HIMU, EM1, and EM2 from global trends between radiogenic isotopes and major elements in ocean island basalts. *Earth and Planetary Science Letters* 276: 175–186.

liquids react with the wall rock (i.e., ambient mantle peridotites) to form hybrid lithologies such as pyroxenites (Yaxley and Green, 1998). In a second step, these hybrid lithologies can produce OIB-like liquids as they melt (they do so at lower pressures than eclogites themselves) (Kogiso and Hirschmann, 2006). Oxide ratios in olivine phenocrysts of Hawaiian lavas have indeed been interpreted as evidence for such a two-stage melting process (Sobolev et al., 2005; 2007). Alternatively, incomplete reaction of eclogitic melts with the wall rock (e.g., along melt channels) can leave a modified liquid that satisfies a range of OIB compositions (Mallik and Dasgupta, 2012). Depending on the extent of wall-rock reaction, either alkalic or tholeiitic liquids are generated.

Hornblendite veins are another viable source lithology for OIB (Pilet et al., 2008; **Figure 12**). Their origin is thought to be a near-ridge environment, where low-degree peridotitic melts freeze into the lithospheric mantle (Pilet et al., 2010). Hornblendite or similar metasomatic lithologies (Halliday et al., 1995; Niu et al., 2002b) have indeed been mapped in peridotite massifs and sampled as xenoliths (Frey et al., 1985; O'Reilly and Griffin, 1988; Takazawa et al., 2000). The metasomatized base of the oceanic lithosphere may contribute to producing OIB-like liquids as it is moved over a hotspot or as it is decompressed as part of a mantle upwelling after recycling. In particular, the rather high K_2O/Na_2O and Al_2O_3/MgO ratios in OIB may indicate the presence of hornblendite in the source or at least provide evidence for the equilibration of the rising magmas with hornblendite.

7.10.3.6.3 The origin of geographic variations in OIB geochemistry

The variability of hotspot geochemical signatures has typically been ascribed to mantle source heterogeneity. In the framework of plume theory, geographic patterns in OIB geochemistry, such as the large variability of isotopic signatures among South Pacific and Darwin Rise OIB (Koppers et al., 2003; Staudigel et al., 1991) or the distinction between the 'Loa' and the 'Kea' trends of Hawaiian volcanism (Abouchami et al., 2005), are thought to reflect heterogeneity in the deep mantle (Farnetani and Hofmann, 2009; Huang et al., 2011; Konter et al., 2008; Koppers and Watts, 2010; Weis et al., 2011). For example, deep mantle heterogeneity is expected to be directly reflected as lateral zonation of the plume conduit (Farnetani and Hofmann, 2009, 2010; Farnetani et al., 2012; **Figure 26(b)**). However, as major-element variations associated with such compositional zonation can severely affect the dynamics of the rising plume (see **Section 7.10.3.3.3**), mapping the deep mantle from geographic patterns of lava compositions is not straightforward (cf. Ballmer et al., in press). Moreover, plume pulsations in general and those induced by solitary waves (e.g., Whitehead and Helfrich, 1988) in particular may tend to homogenize material within the hottest parts of the plume.

Another important aspect to consider is the melting behavior of a heterogeneous source (Phipps Morgan, 1999). Geochemical evidence indicates that such source heterogeneity is present over a range of spatial scales, including scales much smaller than the size of upper mantle melting zones (e.g., Ellam and Stuart, 2004; Ingle et al., 2010; Jackson et al., 2012; Kogiso et al., 2004; Koornneef et al., 2012; Marske et al., 2007; Niu et al., 1996; Phipps Morgan, 1999; Reiners, 2002; Saal et al., 1998; Salters and Dick, 2002; Stracke et al., 2003). As different materials melt over different depth ranges for a given mantle temperature (cf. **Figure 26(a)**), spatial differences in mantle temperature, lithospheric thickness, and the rate of mantle flow through the melting zone can influence the proportions, to which the materials are represented in the lavas. For example, MOR melting is expected to be dominated by the peridotitic DM matrix, a refractory yet the most abundant component, because the thin lithosphere at and near MORs allows for the greatest amount of decompression. Mainly due to the effects of a thicker lithosphere, hotspot magmatism away from MORs is instead predicted to be more heavily influenced by the less abundant but more fertile components, and thus contribute to some of the geochemical differences between OIB and MORB (Ito and Mahoney, 2005a; 2006). In such a framework, mechanisms for non-hot spot volcanism (see **Sections 7.10.3.3.5–7.10.3.3.8**) occurring well away from MORs would similarly generate OIB-like liquids (e.g., Ballmer et al., 2010). First-order global trends indeed display significant correlations of major and trace-element compositions of OIB with the thickness of the lithosphere at the time of volcanism (Dasgupta et al., 2010; Humphreys and Niu, 2009; Ito and Mahoney, 2005b; Keller et al., 2004).

At the regional scale of an individual island group, lateral gradients in temperature and mantle upwelling rate can also give rise to systematic geographic variations in lava chemistry, independent of any lateral compositional zonation of the mantle source (Bianco et al., 2008, 2011b, 2013; Ballmer et al., 2010, 2011; Shorttle et al., 2010; Shorttle and Maclennan, 2011; Ballmer et al. 2013a). For example, lateral thermal gradient beneath the Hawaiian hotspot will tend to shift the melting zones of fertile and of refractory mantle components relative to each other to contribute to the apparent 'Loa' versus 'Kea' geochemical trends (cf. Ballmer et al., in press).

7.10.3.7 Plumes in the Context of Mantle Convection and Plate Tectonics

Our discussion of plumes and alternative forms of mantle upwelling is set amid the background of larger-scale processes of plate tectonics and mantle convection. The cooling of the oceanic lithosphere ultimately drives mantle convection and plate motion. While the volume of mantle that participates in the plate tectonic cycle is still debated, a consensus model has emerged of moderated whole mantle convection, with significant material exchange across the 660 km discontinuity. Plumes naturally develop from the hot thermal boundary layer at the CMB and are displaced when rising through the convecting mantle (Husson and Conrad, 2012; Olson and Singer, 1985; Richards and Griffiths, 1989; Steinberger, 2000; Whitehead and Luther, 1975).

7.10.3.7.1 Connection of hotspots with thermochemical structures in the deep mantle

An important line of evidence for mantle plume theory is the connection of seismic structures in the deep mantle with occurrences of hotspot volcanism at the surface (Castillo, 1988; Williams et al., 1998). The locations of active hotspots, as

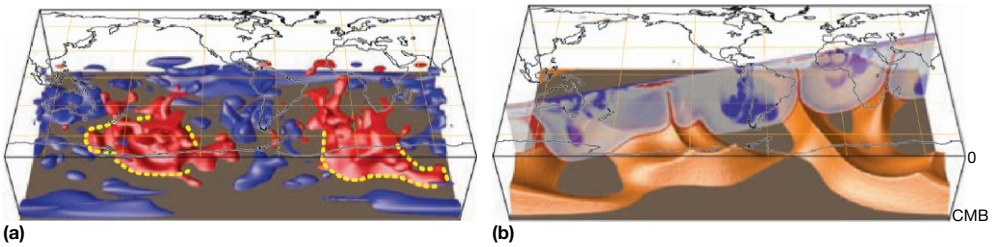


Figure 28 From Garnero and McNamara (2008). Comparison of (a) tomographic shear-wave model S20RTS (Ritsema et al., 2004) with (b) a numerical simulation of whole mantle thermochemical convection. Blue and red isosurfaces in (a) denote lower mantle domains that are anomalously fast and slow (±0.6%), respectively. The large, seismically slow regions at the base of the mantle are the LLSVPs. Yellow dashed lines demarcate the steep gradient in shear-wave velocity as evident along the edges of both LLSVPs (Garnero et al., 2007). (b) Modeled chemically dense material (orange isosurfaces) accumulates as piles at the base of the mantle, a possible explanation for the LLSVPs. In cross section, colors denote temperature from red (hot) to blue (cold).

well as reconstructed locations of extinct hotspots and LIPs, have been shown to be spatially related to the margins of LLSVPs in the lowermost mantle (Burke and Torsvik, 2004; Torsvik et al., 2006, 2010a; **Figures 28(a)** and **29(b)**). This relation holds if the advection of the putative underlying plumes by ambient mantle flow is accounted for (Boschi et al., 2007). The LLSVPs are hot domes or ridges (**Figure 28(b)**) of chemically dense materials (one located beneath Africa, another beneath the South Pacific) with significant topography (500–1000 km) (Garnero and McNamara, 2008; Ishii and Tromp, 1999; Masters et al., 2000; Mosca et al., 2012). Geographic variations seen in OIB geochemistry such as the large-scale DUPAL anomaly (Dupre and Allegre, 1983) have indeed been associated with the LLSVPs (Castillo, 1988; Wen, 2006).

The origin of these large chemically distinct structures is related to the accretion and chemical differentiation of our planet. For example, they may represent primordial material that is unmodified by subsequent melting, or they may be the residuum of a basal magma ocean as formed very early in the Earth's history (Deschamps et al., 2012; Kellogg et al., 1999; Labrosse et al., 2007; Mukhopadhyay, 2012). Along these lines, the LLSVPs have been identified as good candidates to host the FOZO reservoir, a common member in many hotspot lavas (see **Section 7.10.3.6.1**). As an alternative, or perhaps in addition to these ancient origins for the LLSVPs, subducted crust composed of dense eclogitic material is expected to settle and accumulate at the base of the mantle (Brandenburg and van Keken, 2007; Brandenburg et al., 2008; Nakagawa et al., 2010). Such a replenishment of the LLSVPs would further strengthen the connection of mantle plumes with LLSVPs, as subduction-related isotope geochemical signatures are common in OIB (e.g., Hofmann, 1997). Both numerical and analogue models indicate that mantle plumes can indeed be anchored by the stable topography provided by high-density layers (such as by the LLSVPs) (Davaille et al., 2002; Jellinek and Manga, 2002; McNamara and Zhong, 2004).

Global mantle convection models (e.g., McNamara and Zhong, 2005) indicate that any dense thermochemical material would be swept away from subduction zones and toward regions of diffuse upwelling. Using a realistic subduction history and assuming a compositional density difference of ~2%, the dense layer is shaped into structures similar to the imaged LLSVPs (e.g., Steinberger and Torsvik, 2012; **Figure 29**). The predicted diffuse upwellings forming above the LLSVPs may account for the broad dynamic uplift of the overlying lithosphere (i.e., the South Pacific and African Superswells (Adam et al., 2010; Cadio et al., 2011; Hillier and Watts, 2004; Nyblade and Robinson, 1994).

The details of the organization of such upwelling flow, however, remain to be understood (see also **Chapter 7.11**). Many models predict upward flow to focus near the crests of the thermochemical structures (Davaille, 1999; McNamara et al., 2010). Such a flow pattern could only be consistent with the spatial patterns of hotspot volcanism if the LLSVPs represented multiple ridges, swept together by ambient mantle flow to form complex structures (**Figure 28(b)**). However, other models show that plume-like upwelling can also occur from the margins of thermochemical structures. According to Steinberger and Torsvik (2012), large-scale mantle flow tends to sweep the thermal boundary layer toward the margins of the LLSVPs to trigger plume ascent. Plume ascent from the margins is further promoted by a lower compressibility in LLSVP than in ambient mantle materials, which causes LLSVPs to form dome-like structures with steep sides, instead of ridgelike features with shallower sloping sides (Tan and Gurnis, 2007; Tan et al., 2011). Seismic constraints are insufficient to distinguish between dome-like and ridgelike geometries for the LLSVPs (cf. Bull et al., 2009; Lassak et al., 2007; Lay and Garnero, 2011).

7.10.3.7.2 Interaction of plumes with plate tectonic processes

Not only do plate tectonic processes such as slab subduction influence plume locations by piling up thermochemical material at the base of the mantle (**Figures 28** and 29), but also they can directly interact with individual rising plumes. Recent analogue models (Druken, 2012; Kincaid et al., 2013) show that the regional flow field imposed by a sinking slab (cf. Long and Silver, 2009) can distort plume ascent, thereby strongly altering their surface expressions (**Figure 30**). For example, plume flow is predicted to be guided around a slab edge, as is evident, for example, by Samoan-plume geochemical signatures in the juxtaposed northern Lau back-arc basin (e.g., Jackson et al., 2010; Lupton et al., 2012; Lytle et al., 2012). Such a massive disruption of plume ascent could perhaps also give an explanation for the occurrence of extensive secondary volcanism on the Samoan Islands (cf. Hart et al., 2004; Konter and Jackson, 2012; Koppers et al., 2011b; Natland, 1980).

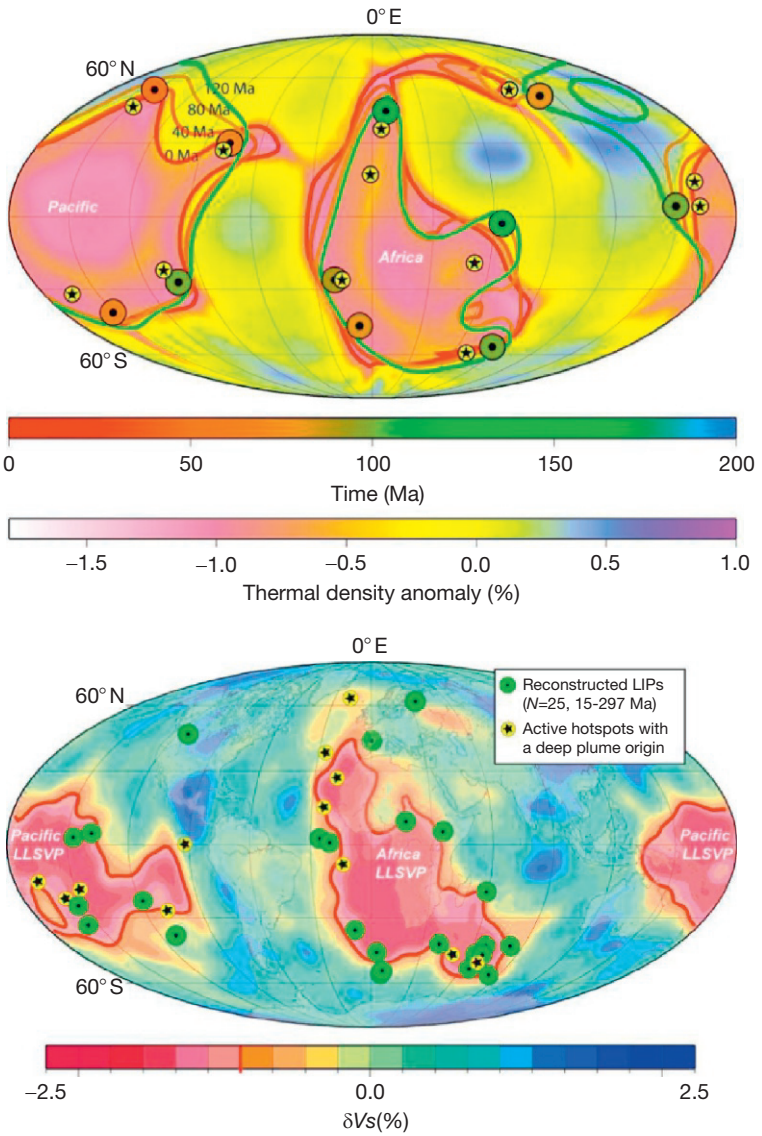


Figure 29 Reproduced from Steinberger B and Antretter M (2006) Conduit diameter and buoyant rising speed of mantle plumes: Implications for the motion of hotspots and shape of plume conduits. *Geochemistry, Geophysics, Geosystems* 7. (top) Representation of their numerical model predictions at 2800 km depth with colors showing thermal density variations (at 0 Ma, i.e., present day; cf. lower color scale) and lines marking the locations of predicted margins of the thermochemical domes (at 120, 80, 40 and 0 Ma). The models are based on subduction locations and fluxes inferred from global plate reconstructions. Large colored circles denote sites, where predicted plume heads first reach the base of the lithosphere (model time: color-coded; cf. upper color scale), corresponding to predicted LIPs. Small yellow circles with stars denote present-day location of plume tails, corresponding to predicted hotspots. Occasionally, a modeled plume disappears or splits into two; therefore, not all predicted LIPs correspond to exactly one predicted hotspot. (bottom) S-wave tomography model SMEAN (Becker and Boschi, 2002) and reconstructed locations of LIPs and present-day hotspot positions (Torsvik et al., 2010a) are shown for comparison.

Slab–plume interaction is also likely to occur in the western United States. Although this region is well illuminated seismically (see **Section 7.10.2.5.2**), no consensus model has emerged for the origin of the Columbia River basalt (CRB) and YSRP hotspot track. The seismic images (**Figure 17**) have been interpreted as revealing warm mantle exploiting a slab gap in the remnant Farallon slab and rising through it when the gap first formed, approximately coinciding with CRB emplacement (James et al., 2011). This interpretation is supported by geodynamic modeling (Liu and Stegman, 2011, 2012). However, it remains controversial whether the mantle upwelling just arises from return flow through the slab gap driven by a local gradient in dynamic pressure (James et al., 2011; Liu and Stegman, 2011, 2012) or is rather a mantle plume driven by excess buoyancy of hot material (Darold and Humphreys, 2013; Obrebski et al., 2010; Sigloch, 2011). Dismissing a plume origin altogether appears to be at odds with observations of a deep-rooted low-velocity anomaly beneath Yellowstone, as well as with the coinciding upwarping of the 660 km discontinuity (Schmandt et al., 2012; **Figure 17(b)**).

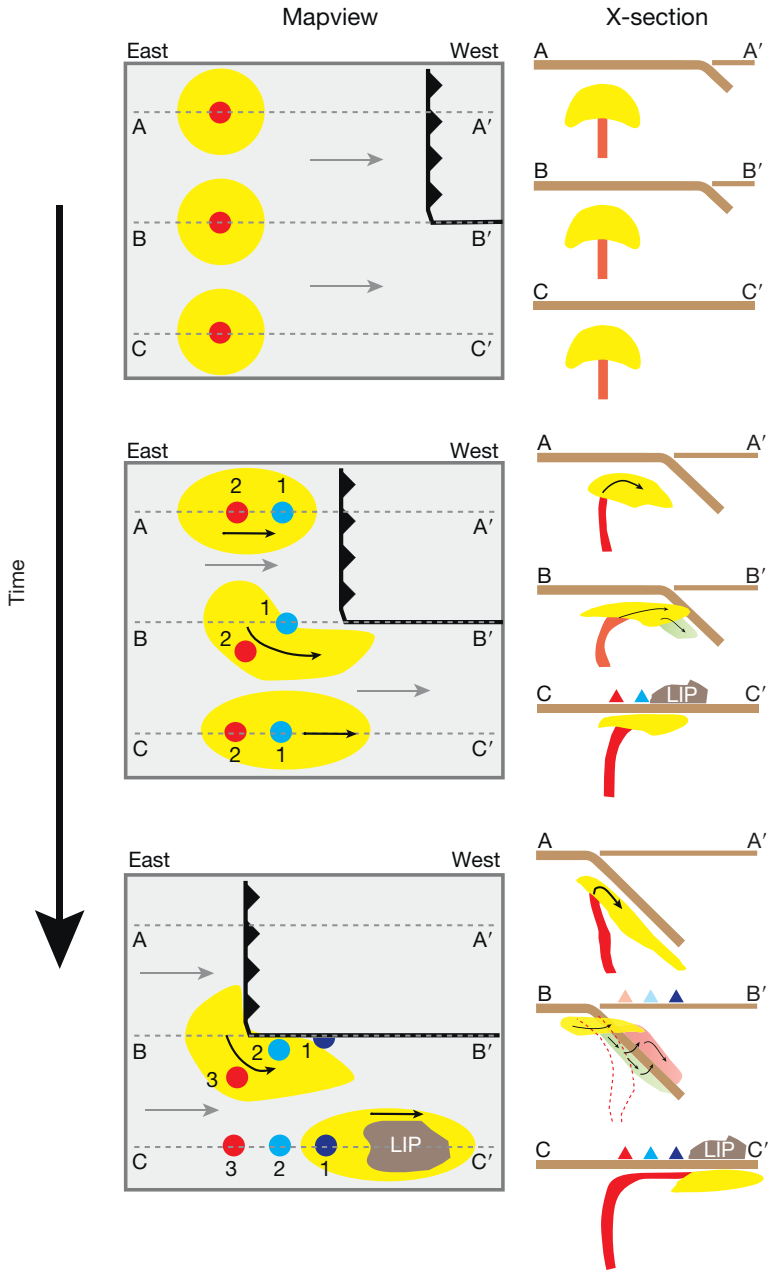


Figure 30 From Druken (2012). Cartoons summarizing the results of analogue models of plume–slab interaction in a time series (top to bottom). In map view, left column shows the surface projection of three upwelling plumes in three distinct tectonic configurations (plume–subducting slab, plume–slab edge, and intraplate plume). Projections of the actively upwelling plume tail (i.e., predicted hotspot location) and plume head are denoted red and yellow, respectively. Blue dots in the bottom panels show locations of previous hotspot activity migrated by plate motion with numbers labeling the progression of volcanism. Right column shows vertical cross sections of the plume head (yellow) and tail (red) along the profiles shown in left column with triangles marking the predicted hotspot track. Material displaced into and out of the plane by complex plume–slab edge interaction in cross section B–B′ is denoted light red and light green, respectively.

The complex seismic structure of the anomaly, which continues to challenge any interpretations in the classical framework of plume theory, may be explained by plume bending around slab fragments and/or plume pulsations rising out of the lower mantle (Schmandt et al., 2012). Moreover, Kincaid et al. (2013) argued that some plume material is systematically entrained into the mantle wedge to explain age-progressive volcanism along the HLPs (**Figure 8**).

Another discussion involves whether segmentation of the Farallon slab occurred independently of (Sigloch, 2011) or was assisted by the putative Yellowstone plume (Obrebski et al., 2010). For example, the arrival of the plume head may have perturbed the slab enough to cause part of the slab to become convectively unstable, drip downward (i.e., the Farallon 'curtain', FC in **Figure 17**), and create the gap through which the plume rose (Darold and Humphreys, 2013). In this

scenario, plume material may have been dammed up by the previously intact slab, with the initial burst of plume material through the slab gap perhaps explaining CRB volcanism.

Slabs are not only able to guide plume ascent but may also trigger mantle upwellings. Small, hydrous plumes have been proposed to rise from the top of the subducted slab to support arc volcanism (Gerya and Yuen, 2003), as well as to hydrate the transition zone (Richard and Bercovici, 2009). Also, slabs stagnating at the base of the upper mantle may sweep material to cause plume-like upwelling and volcanism in the back-arc (Faccenna et al., 2010). In turn, plume heads are able to lubricate or even to actively push plates (Cande and Stegman, 2011; van Hinsbergen et al., 2011).

Plumes also interact with divergent plate boundaries. For example, the Iceland plume alters plate geometries by inducing ridge jumps and ridge propagators to pin the spreading center to the hotspot (Brozena and White, 1990; Hardarson et al., 1997; Hey et al., 2010; Mittelstaedt et al., 2011; 2012). Similar interactions may have pinned the Réunion hotspot close to the Central Indian Ridge until ~30 Ma (cf. Tiwari et al., 2007).

In addition, a variety of geochemical and geophysical evidences show that plume material tends to be drawn toward as well as along seafloor spreading centers (e.g., Schilling et al., 1985; Schilling, 1991; Georgen et al., 2001; see also review by Ito et al., 2003). Geodynamic models show that a combination of self-gravitational spreading, sublithospheric topography, and a dynamic pressure low beneath the ridge facilitate this interaction, whereas divergent flow away from the ridge due to seafloor spreading tends to inhibit the interaction (Feighner and Richards, 1995; Ito et al., 1996, 1997; Ribe, 1996; Ribe and Delattre, 1998; Ribe et al., 1995; Sleep, 1996; Yale and Phipps Morgan, 1998). The depleted residue of hotspot melting and the downwelling curtain forming around a plume pancake may choke nearby MOR melting (Singh et al., 2011; Zhou and Dick, 2013).

7.10.3.7.3 Lithospheric controls on volcanism

In addition to their effects on mantle convective processes, the lithospheric plates can control how ascending magmas are focused to form discrete volcanic edifices. The weight of a volcano exerts stresses on the lithosphere that can redirect magma-filled cracks toward a volcano (Müller et al., 2001), but the effects of lithospheric damage (i.e., thermomechanical erosion of magma pathways and/or melting of the wall rock) are essential to stabilize near-vertical melt extraction directly beneath a volcano (Hieronymus and Bercovici, 2001a). Parameterized models of damage-enhanced lithospheric melt permeability predict the formation of chains with realistic volcano spacing, as distal tectonic stresses are perturbed by the load of the volcanic edifices (Hieronymus and Bercovici, 1999; 2000; 2001b). Perturbation by volcanic loading can also explain the volcano patterns in the Hawaiian Islands (Hieronymus and Bercovici, 1999; ten Brink, 1991), which are expressed as a double chain of staggered volcanoes (Jackson et al., 1972).

Other studies have argued that the lithosphere not only acts as a spatial filter for melts derived from the mantle but also may control the occurrence and timing of volcanic activity itself. An important requirement for this hypothesis is that magmas can be quasi-ubiquitously stabilized in the oceanic asthenosphere (Hirschmann, 2010) or at the base of the oceanic plates (Sakamaki et al., 2013). Such a stabilization has been advocated for to explain the seismic and electric properties of the asthenosphere and/or the sharpness of the lithosphere–asthenosphere boundary (Anderson and Sammis, 1970; Kawakatsu et al., 2009; Naif et al., 2013), an interpretation that however remains controversial (Faul and Jackson, 2005; Karato, 2012, 2013a,b; Karato and Jung, 1998; Stixrude and Lithgow-Bertelloni, 2005). Explaining these geophysical properties and intraplate volcanism at the same time by this mechanism, however, requires a mechanism for melt replenishment.

Volcanic ridges or island chains have been specifically related to intraplate deformation in response to regional (e.g., Clouard and Gerbault, 2008; Sandwell et al., 1995) and local (e.g., Mittelstaedt and Ito, 2005) tectonic stresses or thermal contraction (Gans et al., 2003; Sandwell and Fialko, 2004). For example, Clouard and Gerbault (2008) argued that regional transtensional stresses associated with the rapid subduction along the Tonga–Kermadec Trench damage the Pacific Plate to allow ascent of magmas along a belt across French Polynesia. Hieronymus and Bercovici (1999, 2000) had shown that perturbation of the stress field by volcanic loading (see preceding text) controls the propagation direction (i.e., parallel to deviatoric tensile stresses) of stress-induced volcanism with opportunities for rejuvenated volcanism on the second volcano of the self-propagating chain. Propagation speed and extent are predicted to be limited by the growth of the volcanic edifices and the abundance of sublithospheric melts, respectively. For example, the availability of melts (if not ubiquitous) as potentially controlled by large-scale mantle convection (e.g., diffuse upwelling beneath the South Pacific Superswell) or more localized asthenospheric upwelling (see **Sections 7.10.3.3.5–7.10.3.3.8**) may restrict the final length of the chain.

Accordingly, discrimination between lithospheric controls and other mechanisms for non-hot spot intraplate volcanism remains challenging. Whereas plume theory provides relatively clear and unique testable predictions in terms of geographic age patterns, predictions related to lithospheric controls, small-scale convection (**Section 7.10.3.3.5**), buoyant decompression melting (**Section 7.10.3.3.6**), and shear-driven upwelling (**Section 7.10.3.3.7**) are generally more complex. Distinguishing between these mechanisms thus requires sufficient geochronological data (and appropriate statistical analysis) and/or additional geophysical constraints (Harmon et al., 2011). Future such efforts are needed to understand whether it is melt production, extraction, or ascent that is the bottleneck that limits volcanic activity in the interior of oceanic as well as continental plates.

7.10.4 Conclusion

In the past few decades, our understanding of hotspots and melting anomalies has seen great advancements. It has been realized that not all melting anomalies are related to hotspot activity and that not all hotspots are necessarily underlain by a mantle plume. Many alternative mechanisms for sustaining mantle melting have been proposed and in some cases have been successfully tested with observations (see **Sections 7.10.3.3.5–7.10.3.3.8** and **7.10.3.7.3**). In particular, non-hot spot volcano

chains are best explained by non-plume mechanisms. Yet, plumes remain a viable explanation for large and long-lived hotspots and many LIPs.

Classical plume theory makes a range of specific testable predictions, some of which are successful, whereas others fail at a given hotspot. While the most robust predictions of plume theory have been confirmed to at least the first order at many hotspot chains, second-order departures of observations from predictions require revision of the classical concept. These departures may be related to, for example, thermochemical convection, variable mantle properties, and interaction of plumes with large-scale or regional mantle flow. That plumes may indeed rise from the lower mantle has been confirmed by the spatial correlation of hotspots with structures in the deep mantle and, for a subset of hotspots, by regional seismic tomography studies. Future advancements to understand mantle dynamics and composition by studying melting anomalies will require interdisciplinary efforts including geophysicists, geochemists, geologists, and mineral physicists.

Acknowledgments

We thank Dave Bercovici for editorial guidance and comments that helped to greatly improve the manuscript. M.D.B. and G.I. have been supported by NSF grant EAR-1141938.

References

Aarnes I, Svensen H, Connolly JAD, and Podladchikov YY (2010) How contact metamorphism can trigger global climate changes: Modeling gas generation around igneous sills in sedimentary basins. *Geochimica et Cosmochimica Acta* 74: 7179–7195.

Abalos B, Fountain DM, Ibarguchi JIG, and Puelles P (2011) Eclogite as a seismic marker in subduction channels: Seismic velocities, anisotropy, and petrofabric of Cabo Ortegal eclogite tectonites (Spain). *Geological Society of America Bulletin* 123: U439–U477.

Abouchami W, Hofmann AW, Galer SJG, Frey FA, Eisele J, and Feigenson M (2005) Lead isotopes reveal bilateral asymmetry and vertical continuity in the Hawaiian mantle plume. *Nature* 434: 851–856.

Acton GD and Gordon RG (1994) Paleomagnetic tests of Pacific plate reconstructions and implications for motion between hotspots. *Science* 263: 1246–1254.

Adam C and Bonneville A (2005) Extent of the South Pacific Superswell. *Journal of Geophysical Research* 110.

Adam C and Bonneville A (2008) No thinning of the lithosphere beneath northern part of the Cook-Austral volcanic chains. *Journal of Geophysical Research* 113.

Adam C, Vidal V, and Escartin J (2007) 80-Myr history of buoyancy and volcanic fluxes along the trails of the Walvis and St. Helena hotspots (South Atlantic). *Earth and Planetary Science Letters* 261: 432–442.

Adam C, Yoshida M, Isse T, Suetsugu D, Fukao Y, and Barruol G (2010) South Pacific hotspot swells dynamically supported by mantle flows. *Geophysical Research Letters* 37.

Adams A, Nyblade A, and Weeraratne D (2012) Upper mantle shear wave velocity structure beneath the East African plateau: Evidence for a deep, plateauwide low velocity anomaly. *Geophysical Journal International* 189: 123–142.

Afonso JC, Zlotnik S, and Fernandez M (2008) Effects of compositional and rheological stratifications on small-scale convection under the oceans: Implications for the thickness of oceanic lithosphere and seafloor flattening. *Geophysical Research Letters* 35.

Agrusta R, Arcay D, Tommasi A, Davaille A, Ribe N, and Gerya T (2013) Small-scale convection in a plume-fed low-viscosity layer beneath a moving plate. *Geophysical Journal International* 194: 591–610.

Albers M and Christensen UR (1996) The excess temperature of plumes rising from the core-mantle boundary. *Geophysical Research Letters* 23: 3567–3570.

Ali JR, Thompson GM, Zhou MF, and Song XY (2005) Emeishan large igneous province, SW China. *Lithos* 79: 475–489.

Allegre CJ, Birck JL, Capmas F, and Courtillot V (1999) Age of the Deccan traps using ^{187}Re–^{187}Os systematics. *Earth and Planetary Science Letters* 170: 197–204.

Allen MB, Anderson L, Searle RC, and Buslov M (2006) Oblique rift geometry of the West Siberian Basin: Tectonic setting for the Siberian flood basalts. *Journal of the Geological Society* 163: 901–904.

Allen RM, Nolet G, Morgan WJ, et al. (1999) The thin hot plume beneath Iceland. *Geophysical Journal International* 137: 51–63.

Allen RM, Nolet G, Morgan WJ, et al. (2002) Imaging the mantle beneath Iceland using integrated seismological techniques. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000595.

Allen RM and Tromp J (2005) Resolution of regional seismic models: Squeezing the Iceland anomaly. *Geophysical Journal International* 161: 373–386.

Anderson DL (1989) *Theory of the Earth*, p. 366. Brookline Village, MA: Blackwell.

Anderson DL (1994a) The sublithospheric mantle as the source of continental flood basalts; the case against the continental lithosphere and plume head reservoirs. *Earth and Planetary Science Letters* 123: 269–280.

Anderson DL (1994b) Superplumes or supercontinents? *Geology* 22: 39–42.

Anderson DL (2005) Large igneous provinces, delamination, and fertile mantle. *Elements* 1: 271–275.

Anderson DL and Sammis C (1970) Partial melting in the upper mantle. *Physics of the Earth and Planetary Interiors* 3: 41–50.

Andrews DL, Gordon RG, and Horner-Johnson BC (2006) Uncertainties in plate reconstructions relative to the hotspots: Pacific-hotspot rotations and uncertainties for the past 68 million years. *Geophysical Journal International* 166: 939–951.

Antretter M, Riisager P, Hall S, Zhao X, and Steinberger B (2004) Modelled paleolatitudes for the Louisville hot spot and the Ontong Java Plateau. *Geological Society of America* 299: 21–30, Special Publication.

Aoki I and Takahashi E (2004) Density for MORB eclogite in the upper mantle. *Physics of the Earth and Planetary Interiors* 143–144: 129–143.

Asaadi N, Ribe NM, and Sobouti F (2011) Inferring nonlinear mantle rheology from the shape of the Hawaiian swell. *Nature* 473: U243–U501.

Asimow P and Langmuir CH (2003) The importance of water to oceanic melting regimes. *Nature* 421: 815–820.

Ayalew D, Barbey P, Marty B, Reisberg L, Yirgu G, and Pik R (2002) Source, genesis, and timing of giant ignimbrite deposits associated with Ethiopian continental flood basalts. *Geochimica et Cosmochimica Acta* 66: 1429–1448.

Bagley B and Nyblade AA (2013) Seismic anisotropy in eastern Africa, mantle flow, and the African superplume. *Geophysical Research Letters* 40: 1500–1505.

Ballmer MD, Conrad CP, Smith EI, and Harmon N (2013a) Non-hotspot volcano chains produced by migration of shear-driven upwelling toward the East Pacific Rise. *Geology* 371: 199–202.

Ballmer MD, Ito G, and Cheng C (in press) Thermochemical convection of bilaterally asymmetric plumes and implications for Hawaiian lava composition. In: Carey R, Poland M, Cayol V, and Weis D (eds.) *Hawaiian Volcanism, From Source to Surface. AGU Monograph* 208.

Ballmer MD, Ito G, van Hunen J, and Tackley PJ (2010) Small-scale sublithospheric convection reconciles geochemistry and geochronology of 'Superplume' volcanism in the western and south Pacific. *Earth and Planetary Science Letters* 290: 224–232.

Ballmer MD, Ito G, van Hunen J, and Tackley PJ (2011) Spatial and temporal variability in Hawaiian hotspot volcanism induced by small-scale convection. *Nature Geoscience* 4: 457–460.

Ballmer MD, Ito G, Wolfe CJ, and Solomon SC (2013b) Double layering of a thermochemical plume in the upper mantle beneath Hawaii. *Earth and Planetary Science Letters* 376: 155–164.

Ballmer MD, van Hunen J, Ito G, Bianco TA, and Tackley PJ (2009) Intraplate volcanism with complex age-distance patterns: A case for small-scale sublithospheric convection. *Geochemistry, Geophysics, Geosystems* 10.

Ballmer MD, van Hunen J, Ito G, Tackley PJ, and Bianco TA (2007) Non-hotspot volcano chains originating from small-scale sublithospheric convection. *Geophysical Research Letters* 34.

Barry TL, Self S, Kelley SP, Reidel S, Hooper P, and Widdowson M (2010) New Ar-40/Ar-39 dating of the Grande Ronde lavas, Columbia River Basalts, USA: Implications for duration of flood basalt eruption episodes. *Lithos* 118: 213–222.

Bastow ID, Stuart GW, Kendall JM, and Ebinger CJ (2005) Upper-mantle seismic structure in a region of incipient continental breakup: Northern Ethiopian rift. *Geophysical Journal International* 162: 479–493.

Becker TW and Boschi L (2002) A comparison of tomographic and geodynamic mantle models. *Geochemistry, Geophysics, Geosystems* 3.

Becker TW, Faccenna C, Humphreys ED, Lowry AR, and Miller MS (2013) Static and dynamic support of western United States topography. *Earth and Planetary Science Letters*. http://dx.doi.org/10.1016/j.epsl.2013.10.012.

Beier C, Haase KM, and Turner SP (2012) Conditions of melting beneath the Azores. *Lithos* 144: 1–11.

Benard F, Callot J-P, Vially R, et al. (2010) The Kerguelen plateau: Records from a long-living/composite microcontinent. *Marine and Petroleum Geology* 27: 633–649.

Benoit MH, Nyblade AA, and VanDecar JC (2006) Upper mantle P-wave speed variations beneath Ethiopia and the origin of the Afar hotspot. *Geology* 34: 329–332.

Bercovici D and Kelly A (1997) The non-linear initiation of diapirs and plume heads. *Physics of the Earth and Planetary Interiors* 101: 119–130.

Bianco TA, Conrad CP, and Smith EI (2011a) Time dependence of intraplate volcanism caused by shear-driven upwelling of low-viscosity regions within the asthenosphere. *Journal of Geophysical Research* 116.

Bianco TA, Ito G, Becker JM, and Garcia MO (2005) Secondary Hawaiian volcanism formed by flexural arch decompression. *Geochemistry, Geophysics, Geosystems* 6: Q08009. http://dx.doi.org/10.01029/02005GC000945.

Bianco TA, Ito G, van Hunen J, Ballmer MD, and Mahoney JJ (2008) Geochemical variation at the Hawaiian hot spot caused by upper mantle dynamics and melting of a heterogeneous plume. *Geochemistry, Geophysics, Geosystems* 9.

Bianco TA, Ito G, van Hunen J, Ballmer MD, and Mahoney JJ (2011b) Geochemical variations at intraplate hot spots caused by variable melting of a veined mantle plume. *Geochemistry, Geophysics, Geosystems* 12.

Bianco TA, Ito G, van Hunen J, Mahoney JJ, and Ballmer MD (2013) Geochemical variations at ridge-centered hotspots caused by variable melting of a veined mantle plume. *Earth and Planetary Science Letters* 371: 191–202.

Biggin AJ, Steinberger B, Aubert J, et al. (2012) Possible links between long-term geomagnetic variations and whole-mantle convection processes. *Nature Geoscience* 5: 526–533.

Bijwaard H and Spakman W (1999) Tomographic evidence for a narrow whole mantle plume below Iceland. *Earth and Planetary Science Letters* 166: 121–126.

Bijwaard H, Spakman W, and Engdahl ER (1998) Closing the gap between regional and global travel time tomography. *Journal of Geophysical Research* 103: 30055–30078.

Bjarnason IT, Silver PG, Rumpker G, and Solomon SC (2002) Shear wave splitting across the Iceland hot spot: Results from the ICEMELT experiment. *Journal of Geophysical Research* 107: 2382. http://dx.doi.org/10.1029/2001JB000916.

Black BA, Elkins-Tanton LT, Rowe MC, and Peate IU (2012) Magnitude and consequences of volatile release from the Siberian Traps. *Earth and Planetary Science Letters* 317: 363–373.

Blackburn TJ, Olsen PE, Bowring SA, et al. (2013) Zircon U–Pb geochronology links the End-Triassic extinction with the Central Atlantic Magmatic Province. *Science* 340: 941–945.

Boehler R (2000) High-pressure experiments and the phase diagram of lower mantle and core materials. *Reviews of Geophysics* 38: 221–245.

Bonatti E and Harrison CGA (1976) Hot lines in the Earth's mantle. *Nature* 263: 402–404.

Bonatti E, Harrison CGA, Fisher DE, et al. (1977a) Easter volcanic chain (southeast Pacific): A mantle hot line. *Journal of Geophysical Research* 82: 2457–2478.

Bonneville A, Dosso L, and Hildenbrand A (2006) Temporal evolution and geochemical variability of the South Pacific superplume activity. *Earth and Planetary Science Letters* 244: 251–269.

Bonneville A, Le Suave R, Audin L, et al. (2002) Arago Seamount: The missing hotspot found in the Austral Islands. *Geology* 30: 1023–1026.

Boschi L, Becker TW, and Steinberger B (2007) Mantle plumes: Dynamic models and seismic images. *Geochemistry, Geophysics, Geosystems* 8.

Boschi L, Becker TW, and Steinberger B (2008) On the statistical significance of correlations between synthetic mantle plumes and tomographic models. *Physics of the Earth and Planetary Interiors* 167: 230–238.

Bourdon B, Joron J-L, Claude-Ivanaj C, and Allegre CJ (1998) U–Th–Pa–Ra systematics for the Grande Comore volcanics: Melting processes in an upwelling plume. *Earth and Planetary Science Letters* 164: 119–133.

Bourdon B, Ribe NM, Stracke A, Saal AE, and Turner SP (2006) Insights into the dynamics of mantle plumes from uranium-series geochemistry. *Nature* 444: 713–717.

Bourdon B, Turner SP, and Ribe NM (2005) Partial melting and upwelling rates beneath the Azores from a U-series isotope perspective. *Earth and Planetary Science Letters* 239: 4256.

Brand U, Posenato R, Came R, et al. (2012) The end-Permian mass extinction: A rapid volcanic CO_2 and CH_4-climatic catastrophe. *Chemical Geology* 322: 121–144.

Brandenburg JP, Hauri EH, van Keken PE, and Ballentine CJ (2008) A multiple-system study of the geochemical evolution of the mantle with force-balanced plates and thermochemical effects. *Earth and Planetary Science Letters* 276: 1–13.

Brandenburg JP and van Keken PE (2007) Deep storage of oceanic crust in a vigorously convecting mantle. *Journal of Geophysical Research* 112.

Breger L, Romanowicz B, and Ng C (2001) The Pacific plume as seen by S, ScS, and SKS. *Geophysical Research Letters* 28: 1859–1862.

Breivik AJ, Mjelde R, Faleide JI, and Murai Y (2006) Rates of continental breakup magmatism and seafloor spreading in the Norway Basin–Iceland plume interaction. *Journal of Geophysical Research – Solid Earth* 111, B07102.

Brozena JM and White RS (1990) Ridge jumps and propagations in the South Atlantic Ocean. *Nature* 348: 149–152.

Brunet D and Yuen DA (2000) Mantle plumes pinched in the transition zone. *Earth and Planetary Science Letters* 178: 13–27.

Bryan SE, Peate IU, Peate DW, et al. (2010) The largest volcanic eruptions on Earth. *Earth-Science Reviews* 102: 207–229.

Buck WR (1986) Small-scale convection induced by passive-rifting: the cause for uplift of rift shoulders. *Earth and Planetary Science Letters* 77: 362–372.

Buck WR and Parmentier EM (1986) Convection beneath young oceanic lithosphere: Implications for thermal structure and gravity. *Journal of Geophysical Research* 91: 1961–1974.

Bull AL, McNamara AK, and Ritsema J (2009) Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters* 278: 152–162.

Burke K (1996) The African plate. *South African Journal of Geology* 99: 339–409.

Burke K and Torsvik TH (2004) Derivation of Large Igneous Provinces of the past 200 million years from long-term heterogeneities in the deep mantle. *Earth and Planetary Science Letters* 227: 531–538.

Cadio C, Ballmer MD, Panet I, Diament M, and Ribe N (2012) New constraints on the origin of the Hawaiian swell from wavelet analysis of the geoid to topography ratio. *Earth and Planetary Science Letters* 359–360: 40–54.

Cadio C, Panet I, Davaille A, Diament M, Metivier L, and de Viron O (2011) Pacific geoid anomalies revisited in light of thermochemical oscillating domes in the lower mantle. *Earth and Planetary Science Letters* 306: 123–135.

Camp VE and Hanan BB (2008) A plume-triggered delamination origin for the Columbia River Basalt Group. *Geosphere* 4: 480–495.

Campbell IH (2007) Testing the plume theory. *Chemical Geology* 241: 153–176.

Campbell IH and Griffiths RW (1990) Implications of mantle plume structure for the evolution of flood basalts. *Earth and Planetary Science Letters* 99: 79–93.

Campbell IH and Kerr AC (2007) The Great Plume Debate: Testing the plume theory. *Chemical Geology* 241: 149–152.

Cande SC and Stegman DR (2011) Indian and African plate motions driven by the push force of the Reunion plume head. *Nature* 475: 47–52.

Cao Q, van der Hilst RD, de Hoop MV, and Shim SH (2011) Seismic imaging of transition zone discontinuities suggests hot mantle west of Hawaii. *Science* 332: 1068–1071.

Caress DW, McNutt MK, Detrick RS, and Mutter JC (1995) Seismic imaging of hotspot-related crustal underplating beneath the Marquesas Islands. *Nature* 373: 600–603.

Carlson RW, Czamanske G, Fedorenko V, and Ilupin I (2006) A comparison of Siberian meimechites and kimberlites: Implications for the source of high-Mg alkalic magmas and flood basalts. *Geochemistry, Geophysics, Geosystems* 7.

Castillo P (1988) The dupal anomaly as a trace of the upwelling lower mantle. *Nature* 336: 667–670.

Chandler MT, Wessel P, Taylor B, Seton M, Kim S-S, and Hyeong K (2012) Reconstructing Ontong Java Nui: Implications for Pacific absolute plate motion, hotspot drift and true polar wander. *Earth and Planetary Science Letters* 331: 140–151.

Chang HK, Kowsmann RO, Figueiredo AMF, and Bender AA (1992) Tectonics and stratigraphy of the East Brazil Rift system – An overview. *Tectonophysics* 213: 97–138.

Chang S-J and Van der Lee S (2011) Mantle plumes and associated flow beneath Arabia and East Africa. *Earth and Planetary Science Letters* 302: 448–454.

Chauvel C, Maury RC, Blais S, et al. (2012) The size of plume heterogeneities constrained by Marquesas isotopic stripes. *Geochemistry, Geophysics, Geosystems* 13.

Chazey WJ III and Neal CR (2004) Large igneous province magma petrogenesis from source to surface: Platinum-group element evidence from Ontong Java Plateau basalts recovered during ODP legs 130 and 192. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and Evolution of the Ontong Java Plateau*, pp. 219–238. London: Geological Society, Special Publications 229.

Chenet A-L, Quidelleur X, Fluteau F, Courtillot V, and Bajpai S (2007) K-40–Ar-40 dating of the Main Deccan large igneous province: Further evidence of KTB age and short duration. *Earth and Planetary Science Letters* 263: 1–15.

Cheng C, Allen RM, Porritt RW, Ballmer MD (in press) Seismic constrains on a double-layered asymmetric whole-mantle plume beneath Hawaii. In Carey R, Poland M, Cayol V, Weis D (eds.) *Hawaiian Volcanism, From Source to Surface, AGU Monograph* 208.

Christensen U (1984) Instability of a hot boundary-layer and initiation of thermo-chemical plumes. *Annales Geophysicae* 2: 311–319.

Christensen UR and Hofmann AW (1994) Segregation of subducted oceanic crust and the convecting mantle. *Journal of Geophysical Research* 99: 19867–19884.

Clague DA and Dalrymple GB (1987) The Hawaiian-Emperor volcanic chain. In: Decker RW, Wright TL, and Stauffer PH (eds.) *Volcanism in Hawaii. U.S. Geological Survey Professional Paper*, vol. 1.

Class C, Goldstein SL, Stute M, Kurz MD, and Schlosser P (2005) Grand Comore Island: A well-constrained "low He-3/He-4" mantle plume. *Earth and Planetary Science Letters* 233: 391–409.

Clift PD, Carter A, and Hurford AJ (1998) The erosional and uplift history of NE Atlantic passive margins: Constraints on a passing plume. *Journal of the Geological Society of London* 155: 787–800.

Clift PD and Turner J (1995) Dynamic support by the Icelandic Plume and vertical tectonics of the Northeast Atlantic continental margins. *Journal of Geophysical Research* 100: 24473–24486.

Clouard V and Bonneville A (2001) How many Pacific hotspots are fed by deep-mantle plumes? *Geology* 29: 695–698.

Clouard V and Bonneville A (2005) Ages of Seamounts, islands, and plateaus on the Pacific plate. In: Foulger G, Natland JH, Presnall DC, and Anderson DL (eds.) *Plumes, Plates, and Paradigms*, pp. 71–90. Boulder, CO: GSA, Special Paper 388.

Clouard V, Bonneville A, and Gillot P-Y (2003) The Tarava Seamounts: A newly characterized hotspot chain on the South Pacific Superswell. *Earth and Planetary Science Letters* 207: 117–130.

Clouard V and Gerbault M (2008) Break-up spots: Could the Pacific open as a consequence of plate kinematics? *Earth and Planetary Science Letters* 265: 195–208.

Coble MA and Mahood GA (2012) Initial impingement of the Yellowstone plume located by widespread silicic volcanism contemporaneous with Columbia River flood basalts. *Geology* 40: 655–658.

Coffin MF (1992) Emplacement and subsidence of Indian Ocean plateaus and submarine ridges. In: Duncan RA, Rea DK, Kidd RB, Rad UV, and Weissel JK (eds.) *Synthesis of Results from Scientific Drilling in the Indian Ocean. Geophysical Monograph Series,* 70, pp. 115–125. American Geophysical Union.

Coffin MF and Eldholm O (1993) Scratching the surface: Estimating dimensions of large igneous provinces. *Geology* 21: 515–518.

Coffin MF and Eldholm O (1994) Large igneous provinces: Crustal structure, dimensions, and external consequences. *Reviews of Geophysics* 32: 1–36.

Coffin MF, Pringle MS, Duncan RA, et al. (2002) Kerguelen hotspot magma output since 130 Ma. *Journal of Petrology* 43: 1121–1139.

Cohen RS and O'Nions RK (1982) Identification of recycled continental material in the mantle from Sr, Nd, and Pb isotope investigations. *Earth and Planetary Science Letters* 61: 73–84.

Collier JS, Sansom V, Ishizuka O, Taylor RN, Minshull TA, and Whitmarsh RB (2008) Age of Seychelles–India break-up. *Earth and Planetary Science Letters* 272: 264–277.

Collins JA, Vernon FL, Orcutt JA, and Stephen RA (2002) Upper mantle structure beneath the Hawaiian Swell: Constraints from the ocean seismic network pilot experiment. *Geophysical Research Letters* 29: 1522. http://dx.doi.org/10.1029/2001GL013302.

Coltice N, Bertrand H, Rey P, Jourdan F, Phillips BR, and Ricard Y (2009) Global warming of the mantle beneath continents back to the Archaean. *Gondwana Research* 15: 254–266.

Coltice N, Moreira M, Hernlund J, and Labrosse S (2011) Crystallization of a basal magma ocean recorded by helium and neon. *Earth and Planetary Science Letters* 308: 193–199.

Coltice N, Phillips BR, Bertrand H, Ricard Y, and Rey P (2007) Global warming of the mantle at the origin of flood basalts over supercontinents. *Geology* 35: 391–394.

Coltice N and Ricard Y (1999) Geochemical observations and one layer mantle convection. *Earth and Planetary Science Letters* 174: 125–137.

Condie KC (2001) *Mantle Plumes and Their Record in Earth History,* p. 306. Cambridge, UK: Cambridge University Press.

Connolly JAD and Kerrick DM (2002) Metamorphic controls on seismic velocity of subducted oceanic crust at 100–250 km depth. *Earth and Planetary Science Letters* 204: 61–74.

Conrad CP and Behn MD (2010) Constraints on lithosphere net rotation and asthenospheric viscosity from global mantle flow models and seismic anisotropy. *Geochemistry, Geophysics, Geosystems* 11.

Conrad CP, Bianco TA, Smith EI, and Wessel P (2011) Patterns of intraplate volcanism controlled by asthenospheric shear. *Nature Geoscience* 4: 317–321.

Conrad CP, Wu B, Smith EI, Bianco TA, and Tibbetts A (2010) Shear-driven upwrelling induced by lateral viscosity variations and asthenospheric shear: A mechanism for intraplate volcanism. *Physics of the Earth and Planetary Interiors* 178: 162–175.

Constable S and Heinson G (2004) Hawaiian hot-spot swell structure from seafloor MT sounding. *Tectonophysics* 389: 111–124.

Cordery MJ, Davies GF, and Cambell IH (1997) Genesis of flood basalts from eclogite-bearing mantle plumes. *Journal of Geophysical Research* 102: 20179–20197.

Cornwell DG, Hetenyi G, and Blanchard TD (2011) Mantle transition zone variations beneath the Ethiopian Rift and Afar: Chemical heterogeneity within a hot mantle? *Geophysical Research Letters* 38.

Cottaar S and Romanowicz B (2012) An unusually large ULVZ at the base of the mantle near Hawaii. *Earth and Planetary Science Letters* 355: 213–222.

Courtier AM, Jackson MG, Lawrence JF, et al. (2007) Correlation of seismic and petrologic thermometers suggests deep thermal anomalies beneath hotspots. *Earth and Planetary Science Letters* 264: 308–316.

Courtillot V, Davaille A, Besse J, and Stock J (2003) Three distinct types of hotspots in the Earth's mantle. *Earth and Planetary Science Letters* 205: 295–308.

Courtillot V, Gallet Y, Rocchia R, et al. (2000) Cosmic markers, Ar-40/Ar-39 dating and paleomagnetism of the KT sections in the Anjar Area of the Deccan large igneous province. *Earth and Planetary Science Letters* 182: 137–156.

Courtillot V and Olson P (2007) Mantle plumes link magnetic superchrons to phanerozoic mass depletion events. *Earth and Planetary Science Letters* 260: 495–504.

Courtillot VE and Renne PR (2003) On the ages of flood basalt events. *Comptes Rendus Geoscience* 335: 113–140.

Cox KG (1988) The Karoo province. In: MacDougall JD (ed.) *Continental Flood Basalts*, pp. 239–271. Dordrecht: Kluwer Academic.

Crosby AG and McKenzie D (2009) An analysis of young ocean depth, gravity and global residual topography. *Geophysical Journal International* 178: 1198–1219.

Crough ST (1978) Thermal origin of mid-plate hot-spot swells. *Geophysical Journal of the Royal Astronomical Society* 55: 451–469.

Crough ST (1983) Hotspot swells. *Annual Review of Earth and Planetary Sciences* 11: 165–193.

Crow R, Karlstrom K, Asmerom Y, Schmandt B, Polyak V, and DuFrane SA (2011) Shrinking of the Colorado Plateau via lithospheric mantle erosion: Evidence from Nd and Sr isotopes and geochronology of Neogene basalts. *Geology* 39: 27–30.

Cserepes L, Christensen UR, and Ribe NM (2000) Geoid height versus topography for a plume model of the Hawaiian swell. *Earth and Planetary Science Letters* 178: 29–38.

Cserepes L and Yuen DA (2000) On the possibility of a second kind of mantle plume. *Earth and Planetary Science Letters* 183: 61–71.

Cushman B, Sinton J, Ito G, and Dixon JE (2004) Glass compositions, plume-ridge interaction, and hydrous melting along the Galapagos spreading center, 905 degrees W to 98 degrees W. *Geochemistry, Geophysics, Geosystems* 5: Q08E17. http://dx.doi.org/10.1029/2004GC000709.

d'Acremont E, Leroy S, and Burov EB (2003) Numerical modelling of a mantle plume: The plume head-lithosphere interaction in the formation of an oceanic large igneous province. *Earth and Planetary Science Letters* 206: 379–396.

Darold A and Humphreys E (2013) Upper mantle seismic structure beneath the Pacific Northwest: A plume-triggered delamination origin for the Columbia River flood basalt eruptions. *Earth and Planetary Science Letters* 365: 232–242.

Dasgupta R and Hirschmann MM (2006) Melting in the Earth's deep upper mantle caused by carbon dioxide. *Nature* 440: 659–662.

Dasgupta R, Hirschmann MM, and Smith ND (2007) Partial melting experiments of peridotite CO_2 at 3 GPa and genesis of alkalic ocean island basalts. *Journal of Petrology* 48: 2093–2124.

Dasgupta R, Jackson MG, and Lee C-TA (2010) Major element chemistry of ocean island basalts – Conditions of mantle melting and heterogeneity of mantle source. *Earth and Planetary Science Letters* 289: 377–392.

Davaille A (1999) Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* 402: 756–760.

Davaille A, Girard F, and LeBars M (2002) How to anchor hotspots in a convecting mantle? *Earth and Planetary Science Letters* 203: 621–634.

Davaille A, Gueslin B, Massmeyer A, and Di Giuseppe E (2013) Thermal instabilities in a yield stress fluid: Existence and morphology. *Journal of Non-Newtonian Fluid Mechanics* 193: 144–153.

Davaille A and Jaupart C (1994) Onset of thermal convection in fluids with temperature-dependent viscosity: Application to the oceanic mantle. *Journal of Geophysical Research* 99: 19853–19866.

Davaille A and Vatteville J (2005) On the transient nature of mantle plumes. *Geophysical Research Letters* 32. http://dx.doi.org/10.1029/2005GL023029.

Davies GF (1999) *Dynamic Earth*, p. 458. Cambridge, UK: Cambridge University Press.

Davies GF (2010) Noble gases in the dynamic mantle. *Geochemistry, Geophysics, Geosystems* 11.

Davis AS, Gray LB, Clague DA, and Hein JR (2002) The Line Islands revisited: New $^{40}Ar/^{39}Ar$ geochronologic evidence for episodes of volcanism due to lithospheric extension. *Geochemistry, Geophysics, Geosystems* 3(3). http://dx.doi.org/10.1029/2001GC000190.

Davis FA, Hirschmann MM, and Humayun M (2011) The composition of the incipient partial melt of garnet peridotite at 3 GPa and the origin of OIB. *Earth and Planetary Science Letters* 308: 380–390.

Day JMD, Pearson DG, Macpherson CG, Lowry D, and Carracedo J-C (2009) Pyroxenite-rich mantle formed by recycled oceanic lithosphere: Oxygen-osmium isotope evidence from Canary Island lavas. *Geology* 37: 555–558.

Deenen MHL, Ruhl M, Bonis NR, et al. (2010) A new chronology for the end-Triassic mass extinction. *Earth and Planetary Science Letters* 291: 113–125.

DeLaughter JE, Stein CA, and Stein S (2005) Hotspots: A view from the swells. In: Foulger GR, Natland JH, Presnall DC and Anderson DL (eds.), *Plates, Plumes, and Paradigms*, pp. 257–278. Geological Society of America. Special Paper 388.

Deschamps F, Cobden L, and Tackley PJ (2012) The primitive nature of large low shear-wave velocity provinces. *Earth and Planetary Science Letters* 349: 198–208.

Deschamps F, Kaminski E, and Tackley PJ (2011) A deep mantle origin for the primitive signature of ocean island basalt. *Nature Geoscience* 4: 879–882.

Desonie DL and Duncan RA (1990) The Cobb-Eickelberg seamount chain: Hotspot volcanism with mid-ocean ridge basalt affinity. *Journal of Geophysical Research* 95: 12697–12711.

Detrick RS and Crough ST (1978) Island subsidence, hot spots, and lithospheric thinning. *Journal of Geophysical Research* 83: 1236–1244.

DiVenere V and Kent DV (1999) Are the Pacific and Indo-Atlantic hotspots fixed? Testing the plate circuit through Antarctica. *Earth and Planetary Science Letters* 170: 105–117.

Dodson A, Kennedy BM, and DePaolo DJ (1997) Helium and neon isotopes in the Imnaha Basalt, Columbia River Basalt Group: Evidence for a Yellowstone plume source. *Earth and Planetary Science Letters* 150: 443–451.

Doin MP and Fleitout L (1996) Thermal evolution of the oceanic lithosphere: An alternative view. *Earth and Planetary Science Letters* 142: 121–136.

Doubrovine PV, Steinberger B, and Torsvik TH (2012) Absolute plate motions in a reference frame defined by moving hot spots in the Pacific, Atlantic, and Indian oceans. *Journal of Geophysical Research* 117.

Doucet S, Weis D, Scoates JS, Debaille V, and Giret A (2004) Geochemical and Hf–Pb–Sr–Nd isotopic constraints on the origin of the Amsterdam–St. Paul (Indian Ocean) hotspot basalts. *Earth and Planetary Science Letters* 218: 179–195.

Druken K (2012) *Plumes-Slab Interaction Driven by Rollback Subduction*, PhD Thesis, University of Rhode Island.

Du Z, Vinnik LP, and Foulger GR (2006) Evidence from P-to-S mantle converted waves for a flat "660-km" discontinuity beneath Iceland. *Earth and Planetary Science Letters* 241: 271–280.

Duggen S, Hoernle KA, Hauff F, Klugel A, Bouabdellah M, and Thirlwall MF (2009) Flow of Canary mantle plume material through a subcontinental lithospheric corridor beneath Africa to the Mediterranean. *Geology* 37: 283–286.

Dumoulin C, Choblet G, and Doin MP (2008) Convective interactions between oceanic lithosphere and asthenosphere: Influence of a transform fault. *Earth and Planetary Science Letters* 274: 301–309.

Duncan RA (1978) Geochronology of basalts from the Ninetyeast Ridge and continental dispersion in the Eastern Indian Ocean. *Journal of Volcanology and Geothermal Research* 4: 283–305.

Duncan RA (1984) Age progressive volcanism in the New England Seamounts and the opening of the Central Atlantic Ocean. *Journal of Geophysical Research* 89: 9980–9990.

Duncan RA (1991) Age distribution of volcanism along aseismic ridges in the eastern Indian Ocean. *Proceedings of the Ocean Drilling Program Scientific Results* 121, pp. 507–517. College Station, TX: Ocean Drilling Program.

Duncan RA (2002) A time frame for construction of the Kerguelen Plateau and Broken Ridge. *Journal of Petrology* 43: 1109–1119.

Duncan RA, Backman J, Peterson LC, et al. (1990) The volcanic record of the Reunion hotspot. *Proceedings of the Ocean Drilling Program – Scientific Results* 115: 3–10.

Duncan RA and Clague DA (1985) Pacific plate motion recorded by linear volcanic chains. In: Nairn AEM, Stehli FG, and Uyeda S (eds.) *The Ocean Basins and Margins*, pp. 89–121. New York: Plenum Press.

Duncan RA and Hargraves RB (1984) Caribbean region in the mantle reference frame. In: Bonini W, et al. (eds.) *The Caribbean–South American Plate Boundary and Regional Tectonic. Geological Society of America Memoir*, vol. 162, pp. 89–121. Geological Society of America.

Duncan RA and Keller R (2004) Radiometric ages for basement rocks from the Emperor Seamounts, ODP Leg 197. *Geochemistry, Geophysics, Geosystems* 5: Q08L03. http://dx.doi.org/10.1029/2004GC000704.

Duncan RA and Richards MA (1991) Hotspots, mantle plumes, flood basalts, and true polar wander. *Reviews of Geophysics* 29–1: 31–50.

Dupre B and Allegre CJ (1983) Pb–Sr isotope variation in Indian-Ocean basalts and mixing phenomena. *Nature* 303: 142–146.

Dziewonski AM, Lekic V, and Romanowicz BA (2010) Mantle anchor structure: An argument for bottom up tectonics. *Earth and Planetary Science Letters* 299: 69–79.

Ebinger CJ and Sleep NH (1998) Cenozoic magmatism throughout East Africa resulting from impact of a single plume. *Nature* 395: 788–791.

Elkins-Tanton LT (2007) Continental magmatism, volatile recycling, and a heterogeneous mantle caused by lithospheric gravitational instabilities. *Journal of Geophysical Research* 112.

Elkins-Tanton LT and Hager BH (2005) Giant meteoroid impacts can cause volcanism. *Earth and Planetary Science Letters* 239: 219–232.

Ellam RM and Stuart FM (2004) Coherent He–Nd–Sr isotope trends in high $^{3}He/^{4}He$ basalts: Implications for a common reservoir, mantle heterogeneity and convection. *Earth and Planetary Science Letters* 228: 511–523.

Elliot DH and Fleming TH (2004) Occurrence and dispersal of magmas in the Jurassic Ferrar Large igneous province, Antarctica. *Gondwana Research* 7: 223–237.

Emerick CM and Duncan RA (1982) Age progressive volcanism in the Comores Archipelago, western Indian Ocean and implications for Somali plate tectonics. *Earth and Planetary Science Letters* 60: 415–428.

Encarnacion J, Fleming TH, Elliot DH, and Eales HV (1996) Synchronous emplacement of Ferrar and Karoo dolerites and the early breakup of Gondwana. *Geology* 24: 535–538.

Ernst RE (2007) Mafic-ultramafic large igneous provinces (LIPs): Importance of the pre-mesozoic record. *Episodes* 30: 108–114.

Ernst RE and Baragar WRA (1992) Evidence from magnetic fabric for the flow pattern of magma in the Mackenzie Giant Radiating Dyke Swarm. *Nature* 356: 511–513.

Ernst RE and Buchan KL (eds.) (2001) *Mantle Plumes: Their Identification through Time*, p. 593. Boulder, CO: Geological Society of America.

Ernst RE and Buchan KL (2003) Recognizing mantle plumes in the geological record. *Annual Review of Earth and Planetary Sciences* 31: 469–523.

Ewart A, Marsh JS, Milner SC, Duncan AR, Kamber BS, and Armstrong RA (2004) Petrology and geochemistry of Early Cretaceous bimodal continental flood volcanism of the NW Etendeka, Namibia: Part 1, Introduction, mafic lavas and re-evaluation of mantle source components. *Journal of Petrology* 45: 59–105.

Faccenna C, Becker TW, Lallemand S, Lagabrielle Y, Funiciello F, and Piromallo C (2010) Subduction-triggered magmatic pulses: A new class of plumes? *Earth and Planetary Science Letters* 299: 54–68.

Falloon TJ, Danyushevsky LV, Ariskin A, Green DH, and Ford CE (2007) The application of olivine geothermometry to infer crystallization temperatures of parental liquids: Implications for the temperature of MORB magmas. *Chemical Geology* 241: 207–233.

Farnetani CG (1997) Excess temperature of mantle plumes: The role of chemical stratification of D" *Geophysical Research Letters* 24: 1583–1586.

Farnetani CG and Hofmann AW (2009) Dynamics and internal structure of a lower mantle plume conduit. *Earth and Planetary Science Letters* 282: 314–322.

Farnetani CG and Hofmann AW (2010) Dynamics and internal structure of the Hawaiian plume. *Earth and Planetary Science Letters* 295: 231–240.

Farnetani CG, Hofmann AW, and Class C (2012) How double volcanic chains sample geochemical anomalies from the lowermost mantle. *Earth and Planetary Science Letters* 359–360: 240–247.

Farnetani CG, Legras B, and Tackley PJ (2002) Mixing and deformations in mantle plumes. *Earth and Planetary Science Letters* 196: 1–15.

Farnetani CG and Richards MA (1995) Thermal entrainment and melting in mantle plumes. *Earth and Planetary Science Letters* 133: 251–267.

Farnetani CG and Samuel H (2005) Beyond the thermal plume hypothesis. *Geophysical Research Letters* 32: L07311. http://dx.doi.org/10.1029/02005GL022360.

Faul UH and Jackson I (2005) The seismological signature of temperature and grain size variations in the upper mantle. *Earth and Planetary Science Letters* 234: 119–134.

Fee D and Dueker K (2004) Mantle transition zone topography and structure beneath the Yellowstone hotspot. *Geophysical Research Letters* 31: L18603. http://dx.doi.org/10.1029/12004GL020636.

Feigenson MD, Carr MJ, Maharaj SV, Juliano S, and Bolge LL (2004) Lead isotope composition of Central American volcanoes: Influence of the Galapagos plume. *Geochemistry, Geophysics, Geosystems* 5: Q06001. http://dx.doi.org/10.1029/02003GC000621.

Feighner MA and Richards MA (1995) The fluid dynamics of plume-ridge and plume-plate interactions: An experimental investigation. *Earth and Planetary Science Letters* 129: 171–182.
Ferguson DJ, Maclennan J, Bastow ID, et al. (2013) Melting during late-stage rifting in Afar is hot and deep. *Nature* 499: 70–73.
Fitton JG and Godard M (2004) Origin and evolution of magmas on the Ontong Java Plateau. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and Evolution of the Ontong Java Plateau*, pp. 151–178. London: Geological Society, Special Publications 229.
Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) (2004) *Origin and Evolution of the Ontong Java Plateau*, p. 374. London: Geological Society.
Flament N, Rey PF, Coltice N, Dromart G, and Olivier N (2011) Lower crustal flow kept Archean continental flood basalts at sea level. *Geology* 39: 1159–1162.
Fodor RV and Hanan BB (2000) Geochemical evidence for the Trindade hotspot trace: Columbia Seamount ankaramite. *Lithos* 51: 293–304.
Fontaine FR, Hooft EEE, Burkett PG, Toomey DR, Solomon SC, and Silver PG (2005) Shear-wave splitting beneath the Galapagos archipelago. *Geophysical Research Letters* 32.
Forsyth DW, Harmon N, Scheirer DS, and Duncan RA (2006) Distribution of recent volcanism and the morphology of seamounts and ridges in the GLIMPSE study area: Implications for the lithospheric cracking hypothesis for the origin of intraplate, non-hot spot volcanic chains. *Journal of Geophysical Research* 111.
Forte AM, Quere S, Moucha R, et al. (2010) Joint seismic–geodynamic–mineral physical modelling of African geodynamics: A reconciliation of deep-mantle convection with surface geophysical constraints. *Earth and Planetary Science Letters* 295: 329–341.
Fouch MJ (2012) The Yellowstone hotspot: Plume or not? *Geology* 40: 479–480.
Foulger G and Jurdy DM (eds.) (2007) *Plates, Plumes, and Planetary Processes*, p. 997. Geological Society of America.
Foulger G, Natland J, Presnall D, and Anderson DL (eds.) (2006) *Plates, Plumes, and Paradigms,* p. 861. Geological Society of America.
Foulger GR, Pritchard MJ, Julian BR, et al. (2001) Seismic tomography shows that upwelling beneath Iceland is confined to the upper mantle. *Geophysical Journal International* 146: 504–530.
Frey FA, Coffin MF, Wallace PJ, et al. (2000) Origin and evolution of a submarine large igneous province; the Kerguelen Plateau and Broken Ridge, southern Indian Ocean. *Earth and Planetary Science Letters* 176: 73–89.
Frey FA, Suen CJ, and Stockman HW (1985) The Ronda high-temperature peridotite – Geochemistry and petrogenesis. *Geochimica et Cosmochimica Acta* 49: 2469–2491.
Frey FA, Weis D, Borisova AY, and Xu G (2002) Involvement of continental crust in the formation of the Cretaceous Kerguelen Plateau: New perspectives from ODP Leg 120 sites. *Journal of Petrology* 43: 1207–1239.
Gallacher RJ and Bastow ID (2012) The development of magmatism along the Cameroon Volcanic Line: Evidence from teleseismic receiver functions. *Tectonics* 31.
Ganino C and Arndt NT (2009) Climate changes caused by degassing of sediments during the emplacement of large igneous provinces. *Geology* 37: 323–326.
Gans KD, Wilson DS, and Macdonald KC (2003) Pacific Plate gravity lineaments: Diffuse extension or thermal contraction? *Geochemistry, Geophysics, Geosystems* 4: 1074. http://dx.doi.org/10.1029/2002GC000465.
Garcia MO, Swinnard L, Weis D, et al. (2010) Petrology, geochemistry and geochronology of Kaua'i Lavas over 4.5 Myr: Implications for the origin of rejuvenated volcanism and the evolution of the Hawaiian plume. *Journal of Petrology* 51: 1507–1540.
Garnero EJ, Lay T, and McNamara AK (2007) Implications of lower mantle structural heterogeneity for existence and nature of whole mantle plumes. In: Foulger GR and Jurdy DM (eds.) *The Origin of Melting Anomalies: Plates, Plumes and Planetary Processes*, pp. 79–102. Boulder, CO: Geological Society of America, Special Paper 430.
Garnero EJ and McNamara AK (2008) Structure and dynamics of Earth's lower mantle. *Science* 320: 626–628.
Geldmacher J, Hoefig TW, Hauff F, Hoernle K, Garbe-Schoenberg D, and Wilson DS (2013) Influence of the Galapagos hotspot on the East Pacific Rise during Miocene superfast spreading. *Geology* 41: 183–186.
Geldmacher J, Hoernle K, van der Bogaard P, Duggen S, and Werner R (2005) New Ar-40/Ar-39 age and geochemical data from seamounts in the Canary and Madeira volcanic provinces: Support for the mantle plume hypothesis. *Earth and Planetary Science Letters* 237: 85–101.
Gente P, Dyment J, Maia M, and Goslin J (2003) Interaction between the Mid-Atlantic Ridge and the Azores hot spot during the last 85 Myr: Emplacement and rifting of the hot spot-derived plateaus. *Geochemistry, Geophysics, Geosystems* 4: 8514. http://dx.doi.org/10.1029/2003GC000527.
Georgen JE, Lin J, and Dick HJB (2001) Evidence from gravity anomalies for interactions of the Marion and Bouvet hotspots with the Southwest Indian Ridge: Effects of transform offsets. *Earth and Planetary Science Letters* 187: 283–300.
Gerya TV and Yuen DA (2003) Rayleigh–Taylor instabilities from hydration and melting propel "cold plumes" at subduction zones. *Earth and Planetary Science Letters* 212: 47–62.
Ghatak A and Basu AR (2011) Vestiges of the Kerguelen plume in the Sylhet Traps, northeastern India. *Earth and Planetary Science Letters* 308: 52–64.
Gibson SA, Thompson RN, Weska R, Dickin AP, and Leonardos OH (1997) Late Cretaceous rift-related upwelling and melting of the Trindade starting mantle plume head beneath western Brazil. *Contributions to Mineralogy and Petrology* 126: 303–314.
Gill RCO, Holm PM, and Nielsen TFD (1995) Was a short-lived Baffin Bay plume active prior to initiation of the present Icelandic plume? Clues from the high-Mg picrites of West Greenland. *Lithos* 34: 27–39.
Gill RCO, Pedersen AK, and Larsen JG (1992) Tertiary picrites from West Greenland: Melting at the periphery of a plume? In: Storey BC, Alabaster T, and Pankhurst RJ (eds.) *Magmatism and the Causes of Continental Break-up*, pp. 335–348. London: Geological Society, Special Publication 68.
Giordano G, Lucci F, Phillips D, Cozzupoli D, and Runci V (2012) Stratigraphy, geochronology and evolution of the Mt. Melbourne volcanic field (North Victoria Land, Antarctica). *Bulletin of Volcanology* 74: 1985–2005.
Gladczenko TP, Coffin MF, and Eldholm O (1997) Crustal structure of the Ontong Java Plateau: Modeling of new gravity and existing seismic data. *Journal of Geophysical Research* 102: 22711–22729.
Goes S, Spakman W, and Bijwaard H (1999) A lower mantle source for central European volcanism. *Science* 286: 1928–1931.
Gogus OH and Pysklywec RN (2008a) Mantle lithosphere delamination driving plateau uplift and synconvergent extension in eastern Anatolia. *Geology* 36: 723–726.
Gogus OH and Pysklywec RN (2008b) Near-surface diagnostics of dripping or delaminating lithosphere. *Journal of Geophysical Research* 113.
Gonnermann HM and Mukhopadhyay S (2009) Preserving noble gases in a convecting mantle. *Nature* 459: 560–563.
Goodwillie AM (1995) Short-wavelength gravity lineations and unusual flexure results at the Puka Puka volcanic ridge system. *Earth and Planetary Science Letters* 136: 297–314.
Graham DW (2002) Noble gas isotope geochemistry of mid-ocean ridge and ocean island basalts: Characterization of mantle source reservoirs. In: Porcelli D, Ballentine CJ, and Wieler R (eds.) *Noble Gases in Geochemistry and Cosmochemistry. Reviews in Mineralogy and Geochemistry*, vol. 47, pp. 247–318. Washington, DC: Mineralogical Society of America.
Graham DW, Johnson KTM, Priebe LD, and Lupton JE (1999) Hotspot-ridge interaction along the Southeast Indian Ridge near Amsterdam and St. Paul islands: Helium isotope evidence. *Earth and Planetary Science Letters* 167: 297–310.
Grand SP, van der Hilst RD, and Widiyantoro S (1997) Global seismic tomography: A snapshot of convection in the mantle. *Geological Society of America Today* 7: 1–7.
Green DH, Falloon TJ, Eggins SM, and Yaxley GM (2001) Primary magmas and mantle temperatures. *European Journal Of Mineralogy* 13: 437–451.
Griffin TJ and McDougall I (1975) Geochronology of the cainozoic McBride Volcanic Province, Northern Queensland. *Journal of the Geological Society of Australia* 22: 387–396.
Griffiths RW (1986) Thermals in extremely viscous fluids, including the effects of temperature-dependent viscosity. *Journal of Fluid Mechanics* 166: 115–138.
Griffiths RW and Campbell IH (1990) Stirring and structure in mantle starting plumes. *Earth and Planetary Science Letters* 99: 66–78.
Griffiths RW and Campbell IH (1991) On the dynamics of long-lived plume conduits in the convecting mantle. *Earth and Planetary Science Letters* 103: 214–227.
Grigne C, Labrosse S, and Tackley PJ (2007) Convection under a lid of finite conductivity: Heat flux scaling and application to continents. *Journal of Geophysical Research* 112.
Gripp AE and Gordon RG (2002) Young tracks of hotspots and current plate velocities. *Geophysical Journal International* 150: 321–361.
Guillou H, Carracedo JC, Paris R, and Torrado FJP (2004) Implications for the early shield-stage evolution of Tenerife from K/Ar ages and magnetic stratigraphy. *Earth and Planetary Science Letters* 222: 599–614.
Guillou H, Maury RC, Blais S, et al. (2005) Age progression along the Society hotspot chain (French Polynesia) based on new unspiked K–Ar ages. *Bulletin De La Societe Geologique De France* 176: 135–150.

Gurenko AA, Sobolev AV, Hoernle KA, Hauff F, and Schmincke H-U (2009) Enriched, HIMU-type peridotite and depleted recycled pyroxenite in the Canary plume: A mixed-up mantle. *Earth and Planetary Science Letters* 277: 514–524.

Haase KM, Devey CW, and Goldstein SL (1996) Two-way exchange between the Easter mantle plume and the Easter microplate spreading axis. *Nature* 382: 344–346.

Hales TC, Abt DL, Humphreys ED, and Roering JU (2005) A lithospheric instability origin for Columbia River flood basalts and Wallowa Mountains uplift in northeast Oregon. *Nature* 438: 842–845.

Halliday AN, Lee DC, Tommasini S, et al. (1995) Incompatible trace-elements in OIB and MORB and source enrichment in the sub-oceanic mantle. *Earth and Planetary Science Letters* 133: 379–395.

Hames WE, Renne PR, and Ruppel C (2000) New evidence for geologically instantaneous emplacement of earliest Jurassic Central Atlantic magmatic province basalts on the North American margin. *Geology* 28: 859–862.

Hanan BB and Graham DW (1996) Lead and helium isotope evidence from oceanic basalts for a common deep source of mantle plumes. *Science* 272: 991–995.

Hansen SE and Nyblade AA (2013) The deep seismic structure of the Ethiopia/Afar hotspot and the African superplume. *Geophysical Journal International* 194: 118–124.

Hansen SE, Nyblade AA, and Benoit MH (2012) Mantle structure beneath Africa and Arabia from adaptively parameterized P-wave tomography: Implications for the origin of Cenozoic Afro-Arabian tectonism. *Earth and Planetary Science Letters* 319: 23–34.

Hanyu T, Tatsumi Y, Senda R, et al. (2011) Geochemical characteristics and origin of the HIMU reservoir: A possible mantle plume source in the lower mantle. *Geochemistry, Geophysics, Geosystems* 12.

Hardarson BS, Fitton JG, Ellam RM, and Pringle MS (1997) Rift relocation—A geochemical and geochronological investigation of a paleo-rift in northwest Iceland. *Earth and Planetary Science Letters* 153: 181–196.

Hardebol NJ, Pysklywec RN, and Stephenson R (2012) Small-scale convection at a continental back-arc to craton transition: Application to the southern Canadian Cordillera. *Journal of Geophysical Research* 117.

Harmon N, Forsyth DW, and Scheirer DS (2006) Analysis of gravity and topography in the GLIMPSE study region: Isostatic compensation and uplift of the Sojourn and Hotu Matua Ridge systems. *Journal of Geophysical Research* 111.

Harmon N, Forsyth DW, Weeraratne DS, Yang Y, and Webb SC (2011) Mantle heterogeneity and off axis volcanism on young Pacific lithosphere. *Earth and Planetary Science Letters* 311: 306–315.

Harris RN and McNutt MK (2007) Heat flow on hot spot swells: Evidence for fluid flow. *Journal of Geophysical Research* 112.

Hart SR, Coetzee M, Workman RK, et al. (2004) Genesis of the Western Samoa seamount province: Age, geochemical fingerprint and tectonics. *Earth and Planetary Science Letters* 227: 3756.

Hart SR, Hauri EH, Oschmann LA, and Whitehead JA (1992) Mantle plumes and entrainment: Isotopic evidence. *Science* 256: 517–520.

Hart SR, Schilling J-G, and Powell JL (1973) Basalts from Iceland and along the Reykjanes Ridge: Sr isotope geochemistry. *Nature* 246: 104–107.

Hauri E (1996) Major-element variability in the Hawaiian mantle plume. *Nature* 382: 415–419.

Hawkesworth C, Kelley S, Turner S, le Roex A, and Storey B (1999) Mantle processes during Gondwana break-up and dispersal. *Journal of African Earth Sciences* 28: 239–261.

Haxby WF and Weissel JK (1986) Evidence for small-scale mantle convection from Seasat altimeter data. *Journal of Geophysical Research* 91: 3507–3520.

He B, Xu YG, Chung SL, Xiao L, and Wang Y (2003) Sedimentary evidence for a rapid, kilometer-scale crustal doming prior to the eruption of the Emeishan flood basalts. *Earth and Planetary Science Letters* 213: 391–405.

Heintz M, Debayle E, and Vauchez A (2005) Upper mantle structure of the South American continent and neighboring oceans from surface wave tomography. *Tectonophysics* 406: 115–139.

Hekinian R, Stoffers P, and Cheminee J-L (eds.) (2004) *Oceanic Hotspots*. New York: Springer.

Helffrich G (2000) Topography of the transition zone seismic discontinuities. *Reviews of Geophysics* 38: 141–158.

Helffrich G, Faria B, Fonseca JFBD, Lodge A, and Kaneshima S (2010) Transition zone structure under a stationary hot spot: Cape Verde. *Earth and Planetary Science Letters* 289: 156–161.

Helmberger D, Ni SD, Wen LX, and Ritsema J (2000) Seismic evidence for ultralow-velocity zones beneath Africa and eastern Atlantic. *Journal of Geophysical Research* 105: 23865–23878.

Helmberger DV, Wen L, and Ding X (1998) Seismic evidence that the source of the Iceland hotspot lies at the core–mantle boundary. *Nature* 396: 248–251.

Hendrie DB, Kusznir NJ, Morley CK, and Ebinger CJ (1994) Cenozoic extension in Northern Kenya – A quantitative model of rift basin development in the Turkana region. *Tectonophysics* 236: 409–438.

Hernlund JW, Stevenson DJ, and Tackley PJ (2008a) Buoyant melting instabilities beneath extending lithosphere: 2. Linear analysis. *Journal of Geophysical Research* 113.

Hernlund JW, Tackley PJ, and Stevenson DJ (2008b) Buoyant melting instabilities beneath extending lithosphere: 1. Numerical models. *Journal of Geophysical Research* 113.

Herzberg C (2004a) Geodynamic information in peridotite petrology. *Journal of Petrology* 45: 2507–2530.

Herzberg C (2004b) Partial melting below the Ontong Java Plateau. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and Evolution of the Ontong Java Plateau, Special publications 229*, pp. 179–184. London: Geological Society.

Herzberg C (2006) Petrology and thermal structure of the Hawaiian plume from Mauna Kea volcano. *Nature* 444: 605–609.

Herzberg C (2011) Identification of source lithology in the Hawaiian and Canary Islands: Implications for origins. *Journal of Petrology* 52: 113–146.

Herzberg C, Asimow PD, Arndt N, et al. (2007) Temperatures in ambient mantle and plumes: Constraints from basalts, picrites, and komatiites. *Geochemistry, Geophysics, Geosystems* 8.

Herzberg C and Gazel E (2009) Petrological evidence for secular cooling in mantle plumes. *Nature* 458: U619–U683.

Herzberg C and O'Hara MJ (2002) Plume-associated ultramafic magmas of Phanerozoic age. *Journal of Petrology* 43: 1857–1883.

Hey R, Martinez F, Hoskuldsson A, and Benediktsdottir A (2010) Propagating rift model for the V-shaped ridges south of Iceland. *Geochemistry, Geophysics, Geosystems* 11.

Hieronymus CF and Bercovici D (1999) Discrete alternating hotspot islands formed by interaction of magma transport and lithospheric flexure. *Nature* 397: 604–607.

Hieronymus CF and Bercovici D (2000) Non-hotspot formation of volcanic chains: Control of tectonic and flexural stresses on magma transport. *Earth and Planetary Science Letters* 181: 539–554.

Hieronymus CF and Bercovici D (2001a) Focusing of eruptions by fracture wall erosion. *Geophysical Research Letters* 28: 1823–1826.

Hieronymus CF and Bercovici D (2001b) A theoretical model of hotspot volcanism: Control on volcanic spacing and patterns via magma dynamics and lithospheric stresses. *Journal of Geophysical Research* 106: 683–702.

Hillier JK and Watts AB (2004) "Plate-like" subsidence of the East Pacific Rise-South Pacific superswell system. *Journal of Geophysical Research* 109.

Hirano N, Kawamura K, Hattori M, Saito K, and Ogawa Y (2001) A new type of intra-plate volcanism: Young alkali-basalts discovered from the subducting Pacific Plate, northern Japan Trench. *Geophysical Research Letters* 28: 2719–2722.

Hirano N, Koppers AAP, Takahashi A, Fujiwara T, and Nakanishi M (2008) Seamounts, knolls and petit-spot monogenetic volcanoes on the subducting Pacific Plate. *Basin Research* 20: 543–553.

Hirose K, Fei Y, Ma Y, and Ho-Kwang M (1999) The fate of subducted basaltic crust in the Earth's lower mantle. *Nature* 387: 53–56.

Hirose K and Kushiro I (1993) Partial melting of dry peridotites at high-pressures – Determination of compositions of melts segregated from peridotite using aggregates of diamond. *Earth and Planetary Science Letters* 114: 477–489.

Hirschmann MM (2010) Partial melt in the oceanic low velocity zone. *Physics of the Earth and Planetary Interiors* 179: 60–71.

Hirschmann MM, Asimow PD, Ghiorso MS, and Stolper EM (1999) Calculation of peridotite partial melting from thermodynamic models of minerals and melts. III. Controls on isobaric melt production and the effect of water on melt production. *Journal of Petrology* 40: 831–851.

Hirschmann MM, Kogiso T, Baker MB, and Stolper EM (2003) Alkalic magmas generated by partial melting of garnet pyroxenite. *Geology* 31: 481–484.

Hirth G and Kohlstedt DL (1996) Water in the oceanic upper mantle: Implications for rheology, melt extraction, and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144: 93–108.

Hoernle K, Hauff F, and van den Bogaard P (2004) 70 m.y. history (139–69 Ma) for the Caribbean large igneous province. *Geology* 32: 697–700.

Hoernle K, Hauff F, van den Bogaard P, et al. (2010) Age and geochemistry of volcanic rocks from the Hikurangi and Manihiki oceanic Plateaus. *Geochimica et Cosmochimica Acta* 74: 7196–7219.

Hoernle K, van den Bogaard P, Werner R, et al. (2002) Missing history (16–71 Ma) of the Galapagos hotspot: Implications for the tectonic and biological evolution of the Americas. *Geology* 30: 795–798.

Hofmann AW (1997) Mantle geochemistry: The message from oceanic volcanism. *Nature* 385: 219–229.
Hofmann AW and White WM (1982) Mantle plumes from ancient oceanic crust. *Earth and Planetary Science Letters* 57: 421–436.
Hofmann C, Courtillot V, Feraud G, et al. (1997) Timing of the Ethiopian flood basalt event and implications for plume birth and global change. *Nature* 389: 838–841.
Hofmann C, Feraud G, and Courtillot V (2000) Ar-40/Ar-39 dating of mineral separates and whole rocks from the Western Ghats lava pile: Further constraints on duration and age of the Deccan traps. *Earth and Planetary Science Letters* 180: 13–27.
Holm PM, Gill RCO, Pedersen AK, et al. (1993) The tertiary picrites of West Greenland: Contributions from "Icelandic" and other sources. *Earth and Planetary Science Letters* 115: 227–244.
Hooft EEE, Toomey DR, and Solomon SC (2003) Anomalously thin transition zone beneath the Galapagos hotspot. *Earth and Planetary Science Letters* 216: 55–64.
Hooper P, Widdowson M, and Kelley S (2010) Tectonic setting and timing of the final Deccan flood basalt eruptions. *Geology* 38: 839–842.
Hooper PR (1997) The Columbia River flood basalt province: Current status. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. *Geophysical Monograph*, 100, pp. 1–27. Washington, DC: AGU.
Hooper PR, Binger GB, and Lees KR (2002) Ages of the Steens and Columbia River flood basalts and their relationship to extension-related calc-alkalic volcanism in eastern Oregon. *Geological Society of America Bulletin* 114: 43–50.
Hopper JR, Dahl-Jensen T, Holbrook WS, et al. (2003) Structure of the SE Greenland margin from seismic reflection and refraction data: Implications for nascent spreading center subsidence and asymmetric crustal accretion during North Atlantic opening. *Journal of Geophysical Research* 108: 2269.
Houseman GA, McKenzie DP, and Molnar P (1981) Convective instability of a thickened boundary-layer and its relevance for the thermal evolution of continental convergent belts. *Journal of Geophysical Research* 86: 6115–6132.
Houser C and Williams Q (2010) Reconciling Pacific 410 and 660 km discontinuity topography, transition zone shear velocity patterns, and mantle phase transitions. *Earth and Planetary Science Letters* 296: 255–266.
Huang J (2008) Controls on entrainment of a dense chemical layer by thermal plumes. *Physics of the Earth and Planetary Interiors* 166: 175–187.
Huang J, Fu R, and Ke F (2010) Influence of gradational compositional layering on plume entrainment and its implication for geochemistry. *Physics of the Earth and Planetary Interiors* 178: 6879.
Huang J, Zhong S, and van Hunen J (2003) Controls on sublithospheric small-scale convection. *Journal of Geophysical Research* 108. http://dx.doi.org/10.1029/2003JB002456.
Huang S, Hall PS, and Jackson MG (2011) Geochemical zoning of volcanic chains associated with Pacific hotspots. *Nature Geoscience* 4: 874–878.
Huckfeldt M, Courtier AM, and Leahy GM (2013) Implications for the origin of Hawaiian volcanism from a converted wave analysis of the mantle transition zone. *Earth and Planetary Science Letters* 373: 194–204.
Huerta AD, Nyblade AA, and Reusch AM (2009) Mantle transition zone structure beneath Kenya and Tanzania: More evidence for a deep-seated thermal upwelling in the mantle. *Geophysical Journal International* 177: 1249–1255.
Humphreys ER and Niu Y (2009) On the composition of ocean island basalts (OIB): The effects of lithospheric thickness variation and mantle metasomatism. *Lithos* 112: 118–136.
Hung SH, Shen Y, and Chiao LY (2004) Imaging seismic velocity structure beneath the Iceland hot spot: A finite frequency approach. *Journal of Geophysical Research* 109: B08305. http://dx.doi.org/10.1029/02003JB002889.
Hunt AC, Parkinson IJ, Harris NBW, Barry TL, Rogers NW, and Yondon M (2012) Cenozoic volcanism on the Hangai Dome, Central Mongolia: Geochemical evidence for changing melt sources and implications for mechanisms of melting. *Journal of Petrology* 53: 19131942.
Husson L and Conrad CP (2012) On the location of hotspots in the framework of mantle convection. *Geophysical Research Letters* 39.
Hwang YK, Ritsema J, van Keken PE, Goes S, and Styles E (2011) Wavefront healing renders deep plumes seismically invisible. *Geophysical Journal International* 187: 273–277.
Ingle S and Coffin MF (2004) Impact origin for the greater Ontong Java Plateau? *Earth and Planetary Science Letters* 218: 123–134.
Ingle S, Ito G, Mahoney JJ, et al. (2010) Mechanisms of geochemical and geophysical variations along the western Galapagos Spreading Center. *Geochemistry, Geophysics, Geosystems* 11.
Irifune T, Sekine T, Ringwood AE, and Hibberson WO (1986) The eclogite-garnetite transformations at high pressure and some geophysical implications. *Earth and Planetary Science Letters* 77: 245–256.
Ishii M and Tromp J (1999) Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle. *Science* 285: 1231–1236.
Ito E and Takahashi E (1989) Postspinel transformations in the system Mg_2SiO_4-Fe_2SiO_4 and some geophysical implications. *Journal of Geophysical Research* 94: 10637–10646.
Ito G and Clift P (1998) Subsidence and growth of Pacific Cretaceous plateaus. *Earth and Planetary Science Letters* 161: 85–100.
Ito G, Lin J, and Gable C (1997) Interaction of mantle plumes and migrating midocean ridges: Implications for the Galapagos plume-ridge system. *Journal of Geophysical Research* 102: 15403–15417.
Ito G, Lin J, and Gable CW (1996) Dynamics of mantle flow and melting at a ridge-centered hotspot: Iceland and the Mid-Atlantic Ridge. *Earth and Planetary Science Letters* 144: 53–74.
Ito G, Lin J, and Graham D (2003) Observational and theoretical studies of the dynamics of mantle plume-mid-ocean ridge interaction. *Reviews of Geophysics* 41: 1017. http://dx.doi.org/10.1029/2002RG000117.
Ito G and Mahoney J (2006) Melting a high $^3He/^4He$ source in a heterogeneous mantle. *Geochemistry, Geophysics, Geosystems* 7: Q05010. http://dx.doi.org/10.1029/02005GC001158.
Ito G and Mahoney JJ (2005a) Flow and melting of a heterogeneous mantle: 1.Method and importance to the geochemistry of ocean island and mid-ocean ridge basalts. *Earth and Planetary Science Letters* 230: 29–46.
Ito G and Mahoney JJ (2005b) Flow and melting of a heterogeneous mantle: 2. Implications for a chemically nonlayered mantle. *Earth and Planetary Science Letters* 230: 47–63.
Ito G, McNutt M, and Gibson RL (1995) Crustal structure of the Tuamotu Plateau, 15°S, and implications for its origin. *Journal of Geophysical Research* 100: 8097–8114.
Ito G, Shen Y, Hirth G, and Wolfe CJ (1999) Mantle flow, melting, and dehydration of the Iceland mantle plume. *Earth and Planetary Science Letters* 165: 81–96.
Ivanov BA and Melosh HJ (2003) Impacts do not initiate volcanic eruptions: Eruptions close to the crater. *Geology* 31: 869–872.
Jackson ED, Silver EA, and Dalrymple GB (1972) Hawaiian-Emperor chain and its relation to Cenozoic circumpacific tectonics. *Geological Society of America Bulletin* 83: 601–618.
Jackson I (ed.) (1998) *The Earth's Mantle, Composition, Structure, and Evolution*, p. 566. Cambridge, UK: Cambridge University Press.
Jackson MG and Dasgupta R (2008) Compositions of HIMU, EM1, and EM2 from global trends between radiogenic isotopes and major elements in ocean island basalts. *Earth and Planetary Science Letters* 276: 175–186.
Jackson MG, Hart SR, Konter JG, et al. (2010) Samoan hot spot track on a "hot spot highway": Implications for mantle plumes and a deep Samoan mantle source. *Geochemistry, Geophysics, Geosystems* 11.
Jackson MG, Weis D, and Huang S (2012) Major element variations in Hawaiian shield lavas: Source features and perspectives from global ocean island basalt (OIB) systematics. *Geochemistry, Geophysics, Geosystems* 13.
James DE, Fouch MJ, Carlson RW, and Roth JB (2011) Slab fragmentation, edge flow and the origin of the Yellowstone hotspot track. *Earth and Planetary Science Letters* 311: 124–135.
Janney PE, Macdougall JD, Natland JH, and Lynch MA (2000) Geochemical evidence from the Pukapuka volcanic ridge system for a shallow enriched mantle domain beneath the South Pacific Superswell. *Earth and Planetary Science Letters* 181: 47–60.
Jellinek AM and Manga M (2002) The influence of the chemical boundary layer on the fixity, spacing and lifetime of mantle plumes. *Nature* 418: 760–763.
Jellinek AM and Manga M (2004) Links between long-lived hot spots, mantle plumes, D", and plate tectonics. *Reviews of Geophysics* 42: RG3002. http://dx.doi.org/10.1029/2003RG000144.
Jha K, Parmentier EM, and Morgan JP (1994) The role of mantle depletion and melt-retention buoyancy in spreading-center segmentation. *Earth and Planetary Science Letters* 125: 221–234.
Jones AP, Price GD, Price NJ, DeCarli PS, and Clegg RA (2002a) Impact induced melting and the development of large igneous provinces. *Earth and Planetary Science Letters* 202: 551–561.
Jones SM, White N, and Maclennan J (2002b) V-shaped ridges around Iceland: Implications for spatial and temporal patterns of mantle convection. *Geochemistry, Geophysics, Geosystems* 3: 1059. http://dx.doi.org/10.1029/2002GC000361.
Jordahl KA, McNutt MK, and Caress DW (2004) Multiple episodes of volcanism in the Southern Austral Islands: Flexural constraints from bathymetry, seismic reflection, and gravity data. *Journal of Geophysical Research* 109.
Jordan TH (1979) Mineralogies, densities and seismic velocities of garnet lherzolites and their geophysical implications. In: Boyd FR and Meyer HOA (eds.) *Mantle*

Sample: Inclusions in Kimberlites and Other Volcanics, pp. 1–13. Washington, DC: American Geophysical Union.

Jourdan F, Feraud G, Bertrand H, et al. (2004) The Karoo triple junction questioned: Evidence from Jurassic and Proterozoic Ar-40/Ar-39 ages and geochemistry of the giant Okavango dyke swarm (Botswana). *Earth and Planetary Science Letters* 222: 989–1006.

Jourdan F, Feraud G, Bertrand H, et al. (2005) Karoo large igneous province: Brevity, origin, and relation to mass extinction questioned by new Ar-40/Ar-39 age data. *Geology* 33: 745–748.

Jourdan F, Feraud G, Bertrand H, Watkeys MK, Kampunzu AB, and Le Gall B (2006) Basement control on dyke distribution in Large Igneous Provinces: Case study of the Karoo triple junction. *Earth and Planetary Science Letters* 241: 307–322.

Jourdan F, Marzoli A, Bertrand H, et al. (2009) Ar-40/Ar-39 ages of CAMP in North America: Implications for the Triassic-Jurassic boundary and the K-40 decay constant bias. *Lithos* 110: 167–180.

Kaminski E and Jaupart C (2003) Laminar starting plumes in high-Prandtl-number fluids. *Journal of Fluid Mechanics* 478: 287–298.

Karato S-i (2012) On the origin of the asthenosphere. *Earth and Planetary Science Letters* 321: 95–103.

Karato S-i (2013a) Does partial melting explain geophysical anomalies? *Physics of the Earth and Planetary Interiors* 228: 300–306.

Karato S-i (2013b) Theory of isotope diffusion in a material with multiple species and its implications for hydrogen-enhanced electrical conductivity in olivine. *Physics of the Earth and Planetary Interiors* 219: 49–54.

Karato S and Jung H (1998) Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth and Planetary Science Letters* 157: 193–207.

Kawakatsu H, Kumar P, Takei Y, et al. (2009) Seismic evidence for sharp lithosphere-asthenosphere boundaries of oceanic plates. *Science* 324: 499–502.

Keller G, Adatte T, Gardin S, Bartolini A, and Bajpai S (2008) Main Deccan volcanism phase ends near the K-T boundary: Evidence from the Krishna-Godavari Basin, SE India. *Earth and Planetary Science Letters* 268: 293–311.

Keller RA, Graham DW, Farley KA, Duncan RA, and Lupton E (2004) Cretaceous-to-recent record of elevated He-3/He-4 along the Hawaiian-Emperor volcanic chain. *Geochemistry, Geophysics, Geosystems* 5.

Kellogg LH, Hager B, and van der Hilst R (1999) Compositional stratification in the deep mantle. *Science* 99: 276–289.

Kellogg LH and King SD (1997) The effect of temperature dependent viscosity on the structure of new plumes in the mantle: Results of a finite element model in a spherical, axisymmetric shell. *Earth and Planetary Science Letters* 148: 13–26.

Kelly A and Bercovici D (1997) The clustering of rising diapirs and plume heads. *Geophysical Research Letters* 24: 201–204.

Kent AJR, Stolper EM, Francis D, Woodhead J, Frei R, and Eiler J (2004) Mantle heterogeneity during the formation of the North Atlantic Igneous Province: Constraints from trace element and Sr-Nd-Os-O isotope systematics of Baffin Island picrites. *Geochemistry, Geophysics, Geosystems* 5: Q11004. http://dx.doi.org/10.1029/12004GC000743.

Kent RW, Pringle MS, Muller RD, Sauders AD, and Ghose NC (2002) $^{40}Ar/^{39}Ar$ geochronology of the Rajmahal basalts, India, and their relationship to the Kerguelen Plateau. *Journal of Petrology* 43: 1141–1153.

Kent W, Saunders AD, Kempton PD, and Ghose NC (1997a) Rajmahal Basalts, Eastern India: Mantle sources and melt distribution at a volcanic rifted margin. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic and Planetary Flood Volcanism*. *Geophysical Monograph*, 100, pp. 114–123, 145–182. Washington, DC: American Geophysical Union.

Keyser M, Ritter JRR, and Jordan M (2002) 3D shear-wave velocity structure of the Eifel Plume, Germany. *Earth and Planetary Science Letters* 203: 59–82.

Kidder DL and Worsley TR (2010) Phanerozoic Large Igneous Provinces (LIPs), HEATT (Haline Euxinic Acidic Thermal Transgression) episodes, and mass extinctions. *Palaeogeography, Palaeoclimatology, Palaeoecology* 295: 162–191.

Kieffer B, Arndt N, Lapierre H, et al. (2004) Flood and shield basalts from Ethiopia: Magmas from the African superswell. *Journal of Petrology* 45: 793–834.

Kincaid C, Druken KA, Griffiths RW, and Stegman DR (2013) Bifurcation of the Yellowstone plume driven by subduction-induced mantle flow. *Nature Geoscience* 6: 395–399.

King S and Anderson DL (1998) Edge-driven convection. *Earth and Planetary Science Letters* 160: 289–296.

King SD (2007) Hotspots and edge-driven convection. *Geology* 35: 223–226.

King SD and Anderson DL (1995) An alternative mechanism of flood basalt formation. *Earth and Planetary Science Letters* 136: 269–279.

King SD and Ritsema J (2000) African hot spot volcanism: Small-scale convection in the upper mantle beneath cratons. *Science* 290: 1137–1140.

Klein EM and Langmuir CH (1987) Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *Journal of Geophysical Research* 92: 8089–8115.

Koch FW, Wiens DA, Nyblade AA, et al. (2012) Upper-mantle anisotropy beneath the Cameroon Volcanic Line and Congo Craton from shear wave splitting measurements. *Geophysical Journal International* 190: 75–86.

Kogiso T and Hirschmann MM (2006) Partial melting experiments of bimineralic eclogite and the role of recycled mafic oceanic crust in the genesis of ocean island basalts. *Earth and Planetary Science Letters* 249: 188–199.

Kogiso T, Hirschmann MM, and Frost DJ (2003) High-pressure partial melting of garnet pyroxenite: Possible mafic lithologies in the source of ocean island basalts. *Earth and Planetary Science Letters* 216: 603–617.

Kogiso T, Hirschmann MM, and Reiners PW (2004) Length scales of mantle heterogeneities and their relationship to ocean island basalt geochemistry. *Geochimica et Cosmochimica Acta* 68: 345–360.

Kokfelt TF, Hoernle K, and Hauff F (2003) Upwelling and melting of the Iceland Plume from radial variation of (super 238) U- (super 230) Th disequilibria in postglacial volcanic rocks. *Earth and Planetary Science Letters* 214: 167–186.

Kono M (1980) Paleomagnetism of DSDP Leg 55 basalts and implications for the tectonics of the Pacific plate. In: Jackson ED (ed.) *Initial Reports of the Deep Sea Drilling Project*, vol. 55, pp. 737–752.

Konter JG, Hanan BB, Blichert-Toft J, Koppers AAP, Plank T, and Staudigel H (2008) One hundred million years of mantle geochemical history suggest the retiring of mantle plumes is premature. *Earth and Planetary Science Letters* 275: 285–295.

Konter JG and Jackson MG (2012) Large volumes of rejuvenated volcanism in Samoa: Evidence supporting a tectonic influence on late-stage volcanism. *Geochemistry, Geophysics, Geosystems* 13.

Koornneef JM, Stracke A, Bourdon B, et al. (2012) Melting of a two-component source beneath Iceland. *Journal of Petrology* 53: 127–157.

Koppers AA, Staudigel H, Pringle MS, and Wijbrans JR (2003) Short-lived and discontinuous intraplate volcanism in the South Pacific: Hot spots or extensional volcanism. *Geochemistry, Geophysics, Geosystems* 4: 1089. http://dx.doi.org/10.1029/2003GC000533.

Koppers AAP, Duncan RA, and Steinberger B (2004) Implications of a nonlinear $^{40}Ar/^{39}Ar$ age progression along the Louisville seamount trail for models of fixed and moving hot spots. *Geochemistry, Geophysics, Geosystems* 5: Q06L02. http://dx.doi.org/10.1029/2003GC000671.

Koppers AAP, Gowen MD, Colwell LE, et al. (2011a) New Ar-40/Ar-39 age progression for the Louisville hot spot trail and implications for inter-hot spot motion. *Geochemistry, Geophysics, Geosystems* 12.

Koppers AAP, Russell JA, Jackson MG, Konter J, Staudigel H, and Hart SR (2008) Samoa reinstated as a primary hotspot trail. *Geology* 36: 435–438.

Koppers AAP, Russell JA, Roberts J, et al. (2011b) Age systematics of two young en echelon Samoan volcanic trails. *Geochemistry, Geophysics, Geosystems* 12.

Koppers AAP and Staudigel H (2005) Asynchronous bends in Pacific seamounts trails: A case for extensional volcanism? *Science* 307: 904–907.

Koppers AAP, Staudigel H, Morgan JP, and Duncan RA (2007) Nonlinear (40)Ar/(39)Ar age systematics along the Gilbert Ridge and Tokelau Seamount Trail and the timing of the Hawaii-Emperor Bend. *Geochemistry, Geophysics, Geosystems* 8.

Koppers AAP, Staudigel H, Wijbrans JR, and Pringle MS (1998) The Magellan Seamount Trail: Implications for Cretaceous hotspot volcanism and absolute Pacific Plate motion. *Earth and Planetary Science Letters* 163: 53–68.

Koppers AAP and Watts AB (2010) Intraplate seamounts as a window into deep Earth processes. *Oceanography* 23: 42–57.

Koppers AAP, Yamazaki T, Geldmacher J, et al. (2012) Limited latitudinal mantle plume motion for the Louisville hotspot. *Nature Geoscience* 5: 911–917.

Korenaga J (2004) Mantle mixing and continental breakup magmatism. *Earth and Planetary Science Letters* 218: 463–473.

Korenaga J (2005) Why did not the Ontong Java Plateau form subaerially? *Earth and Planetary Science Letters* 234: 385–399.

Korenaga J (2011) Velocity-depth ambiguity and the seismic structure of large igneous provinces: A case study from the Ontong Java Plateau. *Geophysical Journal International* 185: 1022–1036.

Korenaga J and Jordan TH (2004) Physics of multiscale convection in Earth's mantle: Evolution of sublithospheric convection. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2003JB002464.

Korenaga J and Kelemen PB (2000) Major element heterogeneity in the mantle source of the North Atlantic igneous province. *Earth and Planetary Science Letters* 184: 251–268.

Korenaga J and Sager WW (2012) Seismic tomography of Shatsky Rise by adaptive importance sampling. *Journal of Geophysical Research* 117.

Krishna KS, Abraham H, Sager WW, et al. (2012) Tectonics of the Ninetyeast Ridge derived from spreading records in adjacent oceanic basins and age constraints of the ridge. *Journal of Geophysical Research* 117.
Kumagai I, Davaille A, and Kurita K (2007) On the fate of thermally buoyant mantle plumes at density interfaces. *Earth and Planetary Science Letters* 254: 180–193.
Kumagai I, Davaille A, Kurita K, and Stutzmann E (2008) Mantle plumes: Thin, fat, successful, or failing? Constraints to explain hot spot volcanism through time and space. *Geophysical Research Letters* 35.
Kuroda J, Ogawa NO, Tanimizu M, et al. (2007) Contemporaneous massive subaerial volcanism and late cretaceous Oceanic Anoxic Event 2. *Earth and Planetary Science Letters* 256: 211–223.
Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* 450: 866–869.
Langmuir CH, Klein EM, and Plank T (1992) Petrological systematics of mid-ocean ridge basalts: Constraints on melt generation beneath ocean ridges. In: Phipps Morgan J, Blackman DK, and Sinton JM (eds.) *Mantle Flow and Melt Generation at Mid-Ocean Ridges*. *Geophysical Monograph*, vol. 71, pp. 183–280. Washington, DC: American Geophysical Union.
Lanyon R, Varne R, and Crawford AJ (1993) Tasmanian Tertiary basalts, the Balleny plume, and opening of the Tasman Sea (Southwest Pacific Ocean). *Geology* 21: 555–558.
Larsen TB and Yuen DA (1997) Ultrafast upwelling bursting through the upper mantle. *Earth and Planetary Science Letters* 146: 393–399.
Larsen TB, Yuen DA, and Storey M (1999) Ultrafast mantle plumes and implications for flood basalt volcanism in the Northern Atlantic region. *Tectonophysics* 311: 31–43.
Laske G, Markee A, Orcutt JA, et al. (2011) Asymmetric shallow mantle structure beneath the Hawaiian Swell-evidence from Rayleigh waves recorded by the PLUME network. *Geophysical Journal International* 187: 1725–1742.
Laske G and Masters G (1997) A global digital map of sediment thickness. *EOS, Transactions American Geophysical Union* 78(Fall Meeting Supplement): F483.
Laske G, Phipps Morgan J, and Orcutt JA (2007) The Hawaiian SWELL pilot experiment-evidence for lithosphere rejuvenation from ocean bottom surface wave data. In: Foulger GR and Jurdy DM (eds.) *Plates, Plumes, and Planetary Processes*, pp. 209–233. Geological Society of America.
Lassak TM, McNamara AK, and Zhong S (2007) Influence of thermochemical piles on topography at Earth's core-mantle boundary. *Earth and Planetary Science Letters* 261: 443–455.
Lawrence JF and Shearer PM (2008) Imaging mantle transition zone thickness with SdS-SS finite-frequency sensitivity kernels. *Geophysical Journal International* 174: 143–158.
Lawver LA and Mueller RD (1994) Iceland hotspot track. *Geology* 22: 311–314.
Lay T and Garnero EJ (2011) Deep mantle seismic modeling and imaging. In: Jeanloz R and Freeman KH (eds.) *Annual Review of Earth and Planetary Sciences,* 39, pp. 91–123. Palo Alto, CA: Annual Reviews.
Le Pourhiet L, Gurnis M, and Saleeby J (2006) Mantle instability beneath the Sierra Nevada mountains in California and Death Valley extension. *Earth and Planetary Science Letters* 251: 104–119.
Lecheminant AN and Heaman LM (1989) Mackenzie igneous events, Canada – Middle Proterozoic hotspot magmatism associated with ocean opening. *Earth and Planetary Science Letters* 96: 38–48.
Lee C-TA, Luffi P, Hoeink T, Li J, Dasgupta R, and Hernlund J (2010) Upside-down differentiation and generation of a 'primordial' lower mantle. *Nature* 463: U102–U930.
Lee C-TA, Luffi P, Plank T, Dalton H, and Leeman WP (2009) Constraints on the depths and temperatures of basaltic magma generation on Earth and other terrestrial planets using new thermobarometers for mafic magmas. *Earth and Planetary Science Letters* 279: 20–33.
Lee CTA, Lenardic A, Cooper CM, Niu F, and Levander A (2005) The role of chemical boundary layers in regulating the thickness of continental and oceanic thermal boundary layers (vol. 230, p. 379, 2005). *Earth and Planetary Science Letters* 234: 297–298.
Legendre C, Maury RC, Blais S, Guillou H, and Cotten J (2006) Atypical hotspot chains: Evidence for a secondary melting zone below the Marquesas (French Polynesia). *Terra Nova* 18: 210–216.
Lei JS and Zhao DP (2006) A new insight into the Hawaiian plume. *Earth and Planetary Science Letters* 241: 438–453.
Lekic V, Cottaar S, Dziewonski A, and Romanowicz B (2012) Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity. *Earth and Planetary Science Letters* 357: 68–77.
Lenardic A and Jellinek AM (2009) Tails of two plume types in one mantle. *Geology* 37: 127–130.
Leng W and Gurnis M (2012) Shape of thermal plumes in a compressible mantle with depth- dependent viscosity. *Geophysical Research Letters* 39.
Leng W and Zhong S (2008) Controls on plume heat flux and plume excess temperature. *Journal of Geophysical Research* 113.
Levander A, Schmandt B, Miller MS, et al. (2011) Continuing Colorado plateau uplift by delamination-style convective lithospheric downwelling. *Nature* 472: U461–U540.
Li A and Detrick RS (2003) Azimuthal anisotropy and phase velocity beneath Iceland: Implication for plume-ridge interaction. *Earth and Planetary Science Letters* 214: 153165.
Li A and Detrick RS (2006) Seismic structure of Iceland from Rayleigh wave inversions and geodynamic implications. *Earth and Planetary Science Letters* 241: 901–912.
Li X, Kind R, Priestley K, et al. (2000) Mapping the Hawaiian plume conduit with converted seismic waves. *Nature* 405: 938–941.
Li X, Kind R, Yuan X, Woelbern I, and Hanka W (2004) Rejuvenation of the lithosphere by the Hawaiian Plume. *Nature* 427: 827–829.
Li XQ, Kind R, and Yuan XH (2003) Seismic study of upper mantle and transition zone beneath hotspots. *Physics of the Earth and Planetary Interiors* 136: 79–92.
Lin S-C and van Keken PE (2006a) Deformation, stirring and material transport in thermochemical plumes. *Geophysical Research Letters* 33.
Lin S-C and van Keken PE (2006b) Dynamics of thermochemical plumes: 1. Plume formation and entrainment of a dense layer. *Geochemistry, Geophysics, Geosystems* 7: Q02006. http://dx.doi.org/10.1029/02005GC001071.
Lin SC, Kuo BY, Chiao LY, and van Keken PE (2005) Thermal plume models and melt generation in East Africa: A dynamic modeling approach. *Earth and Planetary Science Letters* 237: 175–192.
Lin SC and van Keken PE (2005) Multiple volcanic episodes of flood basalts caused by thermochemical mantle plumes. *Nature* 436: 250–252.
Lin SC and van Keken PE (2006c) Dynamics of thermochemical plumes: 2. Complexity of plume structures and its implications for mapping mantle plumes. *Geochemistry, Geophysics, Geosystems* 7: Q03003. http://dx.doi.org/10.1029/02005GC001072.
Lipman PW and Glazner AF (1991) Introduction to middle tertiary cordilleran volcanism – Magma sources and relations to regional tectonics. *Journal of Geophysical Research* 96: 13193–13199.
Lithgow-Bertelloni C, Richards MA, Conrad CP, and Griffiths RW (2001) Plume generation in natural thermal convection at high Rayleigh and Prandtl numbers. *Journal of Fluid Mechanics* 434: 1–21.
Liu L and Stegman DR (2011) Segmentation of the Farallon slab. *Earth and Planetary Science Letters* 311: 1–10.
Liu L and Stegman DR (2012) Origin of Columbia River flood basalt controlled by propagating rupture of the Farallon slab. *Nature* 482: U386–U1508.
Lodge A and Helffrich G (2006) Depleted swell root beneath the Cape Verde Islands. *Geology* 34: 449–452.
Long MD and Silver PG (2009) Mantle flow in subduction systems: The subslab flow field and implications for mantle dynamics. *Journal of Geophysical Research* 114.
Long MD, Till CB, Druken KA, et al. (2012) Mantle dynamics beneath the Pacific Northwest and the generation of voluminous back-arc volcanism. *Geochemistry, Geophysics, Geosystems* 13.
Lupton JE, Arculus RJ, Evans LJ, and Graham DW (2012) Mantle hotspot neon in basalts from the Northwest Lau Back-arc Basin. *Geophysical Research Letters* 39.
Lustrino M and Wilson M (2007) The circum-Mediterranean anorogenic Cenozoic igneous province. *Earth-Science Reviews* 81: 1–65.
Lynch MA (1999) Linear ridge groups: Evidence for tensional cracking in the Pacific Plate. *Journal of Geophysical Research* 104: 29321–29333.
Lytle ML, Kelley KA, Hauri EH, Gill JB, Papia D, and Arculus RJ (2012) Tracing mantle sources and Samoan influence in the northwestern Lau back-arc basin. *Geochemistry, Geophysics, Geosystems* 13.
Mackay LM, Turner J, Jones SM, and White NJ (2005) Cenozoic vertical motions in the Moray Firth Basin associated with initiation of the Iceland Plume. *Tectonics* 24.
Maclennan J, McKenzie D, and Gronvold K (2001) Plume-driven upwelling under central Iceland. *Earth and Planetary Science Letters* 194: 67–82.
Macnab R, Verhoef J, Roest W, and Arkani-Hamed J (1995) New database documents the magnetic character of the Arctic and North Atlantic. *EOS, Transactions of the American Geophysical Union* 76(45): 449.
Mahoney JJ and Coffin MF (eds.) (1997) *Large Igneous Provinces: Continental, Oceanic and Planetary Flood Volcanism,* p. 438. Washington, DC: American Geophysical Union.
Mahoney JJ, Duncan RA, Tejada MLG, Sager WW, and Bralower TJ (2005) Jurassic-Cretaceous boundary age and mid-oceanic-ridge-type mantle source for Shatsky Rise. *Geology* 33: 185–188.
Mahoney JJ, Fitton JG, Wallace PJ, et al. (2001) Proceedings of the Ocean Drilling Program, Initial Reports, 192 (online). http://www-odp.tamu.edu/publications/192IR7192ir.htm.

Mahoney JJ, Jones WB, Frey FA, Salters VJ, Pyle DG, and Davies HL (1995) Geochemical characteristics of lavas from Broken Ridge, the Naturaliste Plateau and southernmost Kerguelen plateau: Cretaceous plateau volcanism in the SE Indian Ocean. *Chemical Geology* 120: 315–345.

Mahoney JJ, Storey M, Duncan RA, Spencer KJ, and Pringle M (1993) Geochemistry and age of the Ontong Java Plateau. In: Pringle M, Sager W, Sliter W, and Stein S (eds.) *The Mesozoic Pacific: Geology, Tectonics, and Volcanism*. *Geophysical Monograph Seriese,* 77, pp. 233–261. Washington, DC: AGU.

Maia M, Ackermand D, Dehghani GA, et al. (2000) The Pacific-Antarctic Ridge-Foundation Hotspot interaction: A case study of a ridge approaching a hotspot. *Marine Geology* 167: 61–84.

Maia M, Hemond C, and Gente P (2001) Contrasted interactions between plume, upper mantle, and lithosphere: Foundation chain case. *Geochemistry, Geophysics, Geosystems* 2.

Mallik A and Dasgupta R (2012) Reaction between MORB-eclogite derived melts and fertile peridotite and generation of ocean island basalts. *Earth and Planetary Science Letters* 329: 97–108.

Manga M (1997) Interactions between mantle diapirs. *Geophysical Research Letters* 24: 1871–1874.

Marks KM and Sandwell DT (1991) Analysis of geoid height versus topography for oceanic plateaus and swells using nonbiased linear-regression. *Journal of Geophysical Research* 96: 8045–8055.

Marske JP, Pietruszka AJ, Weis D, Garcia MO, and Rhodes JM (2007) Rapid passage of a small-scale mantle heterogeneity through the melting regions of Kilauea and Mauna Loa Volcanoes. *Earth and Planetary Science Letters* 259: 34–50.

Marzoli A, Jourdan F, Puffer JH, et al. (2011) Timing and duration of the Central Atlantic magmatic province in the Newark and Culpeper basins, eastern USA. *Lithos* 122: 175–188.

Marzoli A, Piccirillo EM, Renne PR, et al. (2000) The Cameroon volcanic line revisited: Petrogenesis of continental basaltic magmas from lithospheric and asthenospheric mantle sources. *Journal of Petrology* 41: 87–109.

Marzoli A, Renne PR, Piccirillo EM, Ernesto M, Bellieni G, and De Min A (1999) Extensive 200-million-year-old continental flood basalts of the Central Atlantic Magmatic Province. *Science* 284: 616–618.

Massmeyer A, Di Giuseppe E, Davaille A, Rolf T, and Tackley PJ (2013) Numerical simulation of thermal plumes in a Herschel-Bulkley fluid. *Journal of Non-Newtonian Fluid Mechanics* 195: 32–45.

Masters G, Laske G, Bolton H, and Dziewonski A (2000) The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure. In: *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*. *Geophysical Monograph Series,* 117, pp. 63–87. Washington, DC: AGU.

Matyska C, Moser J, and Yuen DA (1994) The potential influence of radiative heat transfer on the formation of megaplumes in the lower mantle. *Earth and Planetary Science Letters* 125: 255–266.

Matyska C and Yuen DA (2005) The importance of radiative heat transfer on superplumes in the lower mantle with the new post-perovskite phase change. *Earth and Planetary Science Letters* 234: 71–81.

Matyska C and Yuen DA (2006) Lower mantle dynamics with the post-perovskite phase change, radiative thermal conductivity, temperature- and depth-dependent viscosity. *Physics of the Earth and Planetary Interiors* 154: 196–207.

Matyska C, Yuen DA, Wentzcovitch RM, and Qzkova H (2011) The impact of variability in the rheological activation parameters on lower-mantle viscosity stratification and its dynamics. *Physics of the Earth and Planetary Interiors* 188: 1–8.

Mayborn KR and Lesher CE (2004) Paleoproterozoic mafic dike swarms of northeast Laurentia: Products of plumes or ambient mantle? *Earth and Planetary Science Letters* 225: 305–317.

McDougall I and Duncan RA (1988) Age progressive volcanism in the Tasmantid Seamounts. *Earth and Planetary Science Letters* 89: 207–220.

McDougall I, Verwoerd W, and Chevallier L (2001) K-Ar geochronology of Marion Island, Southern Ocean. *Geological Magazine* 138: 1–17.

McKenzie D (1984) The generation and compaction of partially molten rock. *Journal of Petrology* 25: 713–765.

McKenzie D and Bickle MJ (1988) The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology* 29: 625–679.

McNamara AK, Garnero EJ, and Rost S (2010) Tracking deep mantle reservoirs with ultra-low velocity zones. *Earth and Planetary Science Letters* 299: 1–9.

McNamara AK and Zhong SJ (2004) The influence of thermochemical convection on the fixity of mantle plumes. *Earth and Planetary Science Letters* 222: 485–500.

McNamara AK and Zhong SJ (2005) Thermochemical structures beneath Africa and the Pacific Ocean. *Nature* 437: 1136–1139.

McNutt M (1988) Thermal and mechanical properties of the Cape Verde Rise. *Journal of Geophysical Research* 93: 2784–2794.

McNutt M and Bonneville A (2000) A shallow, chemical origin for the Marquesas Swell. *Geochemistry, Geophysics, Geosystems* 1.

McNutt MK (1998) Superswells. *Reviews of Geophysics* 36: 211–244.

McNutt MK (2002) Heat flow variations over Hawaiian swell controlled by near-surface processes, not plume properties. *Geophysical Monograph* 128: 364–365.

McNutt MK, Caress DW, Reynolds J, Jordahl KA, and Duncan RA (1997) Failure of plume theory to explain midplate volcanism in the southern Austral Island. *Nature* 389: 479–482.

McNutt MK and Fischer KM (1987) The South Pacific Superswell. In: Keating GH, Fryer P, Batiza R, and Boehlert GW (eds.) *Seamounts, Islands, and Atolls*. *Geophysical Monograph,* 43, Washington, DC: AGU.

McNutt MK and Judge AV (1990) The Superswell and mantle dynamics beneath the South Pacific. *Science* 248: 969–975.

McNutt MK, Winterer EL, Sager WW, Natland JH, and Ito G (1990) The Darwin Rise: A cretaceous superswell? *Geophysical Research Letters* 17: 1101–1104.

Meert JG and Tamrat E (2006) Paleomagnetic evidence for a stationary Marion hotspot: Additional paleomagnetic data from Madagascar. *Gondwana Research* 10: 340–348.

Mege D and Korme T (2004) Dyke swarm emplacement in the Ethiopian large igneous province: Not only a matter of stress. *Journal of Volcanology and Geothermal Research* 132: 283–310.

Meibom A, Sleep NH, Zahnle K, and Anderson DL (2005) Models for noble gasses in mantle geochemistry: Some observations and alternatives. In: *Plumes, Plates, and Paradigms*, pp. 347–363. Geological Society of America, Special paper 388.

Menzies M, Baker J, Chazot G, and Al'Kadasi M (1997) Evolution of the Red Sea volcanic margin, Western Yemen. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. *Geophysical Monograph,* 100, pp. 29–43. Washington, DC: American Geophysical Union.

Mihalffy P, Steinberger B, and Schmeling H (2008) The effect of the large-scale mantle flow field on the Iceland hotspot track. *Tectonophysics* 447: 5–18.

Milelli L, Fourel L, and Jaupart C (2012) A lithospheric instability origin for the Cameroon Volcanic Line. *Earth and Planetary Science Letters* 335: 80–87.

Minor DR and Mukasa SB (1997) Zircon U-Pb and hornblende Ar-40-Ar-39 ages for the Dufek layered mafic intrusion, Antarctica: Implications for the age of the Ferrar large igneous province. *Geochimica et Cosmochimica Acta* 61: 2497–2504.

Minshull TA, Lane CI, Collier JS, and Whitmarsh RB (2008) The relationship between rifting and magmatism in the northeastern Arabian Sea. *Nature Geoscience* 1: 463–467.

Missenard Y and Cadoux A (2012) Can Moroccan Atlas lithospheric thinning and volcanism be induced by Edge-Driven Convection? *Terra Nova* 24: 27–33.

Mittelstaedt E, Ito G, and van Hunen J (2011) Repeat ridge jumps associated with plume-ridge interaction, melt transport, and ridge migration. *Journal of Geophysical Research* 116.

Mittelstaedt E, Soule S, Harpp K, et al. (2012) Multiple expressions of plume-ridge interaction in the Galapagos: Volcanic lineaments and ridge jumps. *Geochemistry, Geophysics, Geosystems* 13.

Mittelstaedt EL and Ito G (2005) Plume-ridge interaction, lithospheric stresses, and origin of near-ridge volcanic lineaments. *Geochemistry, Geophysics, Geosystems* 6: Q06002. http://dx.doi.org/10.1029/02004GC000860.

Mjelde R, Wessel P, and Mueller RD (2010) Global pulsations of intraplate magmatism through the Cenozoic. *Lithosphere* 2: 361–376.

Molnar P and Stock J (1987) Relative motions of hotspots in the Pacific, Atlantic and Indian Oceans since late Cretaceous time. *Nature* 327: 587–591.

Montelli R, Nolet G, Dahlen FA, and Masters G (2006) A catalogue of deep mantle plumes: New results from finite-frequency tomography. *Geochemistry, Geophysics, Geosystems* 7.

Montigny R, Ngounouno I, and Deruelle B (2004) K-Ar ages of magmatic rocks from the Garoua rift: Their place in the frame of the 'Cameroun Line'. *Comptes Rendus Geoscience* 336: 1463–1471.

Moore WB, Schubert G, and Tackley P (1998) Three-dimensional simulations of plume-lithosphere interaction at the Hawaiian swell. *Science* 279: 1008–1011.

Morgan WJ (1971) Convection plumes in the lower mantle. *Nature* 230: 42–43.

Morgan WJ (1972) Plate motions and deep mantle convection. *Geological Society of America Memoir* 132: 7–22.

Morgan WJ (1978) Rodriguez, Darwin, Amsterdam,. . ., A second type of hotspot island. *Journal of Geophysical Research* 83: 5355–5360.

Morgan WJ (1983) Hotspot tracks and the early rifting of the Atlantic. *Tectonophysics* 94: 123–139.

Morley CK, Wescott WA, Stone DM, Harper RM, Wigger ST, and Karanja FM (1992) Tectonic evolution of The Northern Kenyan rift. *Journal of the Geological Society* 149: 333–348.

Mosar J, Lewis G, and Torsvik TH (2002) North Atlantic sea-floor spreading rates; implications for the Tertiary development of inversion structures of the Norwegian-Greenland Sea. *Geological Society of London* 159: 503–515.
Mosca I, Cobden L, Deuss A, Ritsema J, and Trampert J (2012) Seismic and mineralogical structures of the lower mantle from probabilistic tomography. *Journal of Geophysical Research* 117.
Mueller RD, Roest WR, and Royer J-Y (1998) Asymmetric sea-floor spreading caused by ridge-plume interactions. *Nature* 396: 455–459.
Mukhopadhyay S (2012) Early differentiation and volatile accretion recorded in deep-mantle neon and xenon. *Nature* 486: U101–U124.
Müller JR, Ito G, and Martel SJ (2001) Effects of volcano loading on dike propagation in an elastic half-space. *Journal of Geophysical Research* 106: 11101–11113.
Müller RD, Roest WR, Royer J-Y, Gahagan LM, and Sclater JG (1993a) *A digital age map of the ocean floor*. Scripps Institution of Oceanography, Reference Series.
Müller RD, Royer J-Y, and Lawver LA (1993b) Revised plate motions relative to the hotspots from combined Atlantic and Indian Ocean hotspot tracks. *Geology* 21: 275–278.
Mutter JC and Zehnder CM (1988) Deep crustal structure and magmatic processes: The inception of seafloor spreading in the Norwegian-Greenland Sea. In: Moron AC and Parson LM (eds.) *Early Tertiary Volcanism and the Opening of the NE Atlantic 39*, pp. 35–48. Geological Society of America, Special Publication.
Nadin PA, Kusznir NJ, and Cheadle MJ (1997) Early Tertiary plume uplift of the North Sea and Faeroe-Shetland basins. *Earth and Planetary Science Letters* 148: 109–127.
Naif S, Key K, Constable S, and Evans RL (2013) Melt-rich channel observed at the lithosphere-asthenosphere boundary. *Nature* 495: 356–359.
Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2010) The influence of MORB and harzburgite composition on thermo-chemical mantle convection in a 3-D spherical shell with self-consistently calculated mineral physics. *Earth and Planetary Science Letters* 296: 403–412.
Nakakuki T, Yuen DA, and Honda S (1997) The interaction of plumes with the transition zone under continents and oceans. *Earth and Planetary Science Letters* 146: 379–391.
Nakanishi M, Sager WW, and Klaus A (1999) Magnetic lineations within Shatsky Rise, northwest Pacific Ocean: Implications for hot spot-triple junction interaction and oceanic plateau formation. *Journal of Geophysical Research* 104: 7539–7556.
Nataf H-C (2000) Seismic imaging of mantle plumes. *Annual Review of Earth and Planetary Sciences* 28: 391–417.
Nataf HC and VanDecar J (1993) Seismological detection of a mantle plume? *Nature* 364: 115120.
Natland JH (1980) The progression of volcanism in the Samoan linear volcanic chain. *American Journal of Science* 280A: 709–735.
Neal CR, Mahoney JJ, and Chazey WJ (2002) Mantle sources and the highly variable role of continental lithosphere in basalt petrogenesis of the Kerguelen Plateau and Broken Ridge LIP: Results from ODP Leg 183. *Journal of Petrology* 43: 1177–1205.
Neal CR, Mahoney JJ, Kroenke LW, Duncan RA, and Petterson MG (1997) The Ontong Java Plateau. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. *Geophysical Monograph*, vol. 100, pp. 183–216. Washington, DC: American Geophysical Union.
Nelson ST and Tingey DG (1997) Time-transgressive and extension-related basaltic volcanism in southwest Utah and vicinity. *Geological Society of America Bulletin* 109: 1249–1265.
Ni SD, Helmberger DV, and Tromp J (2005) Three-dimensional structure of the African superplume from waveform modelling. *Geophysical Journal International* 161: 283–294.
Ni SD, Tan E, Gurnis M, and Helmberger D (2002) Sharp sides to the African superplume. *Science* 296: 1850–1852.
Nicolaysen K, Frey FA, Hodges KV, Weis D, and Giret A (2000) $^{40}Ar/^{39}Ar$ geochronology of flood basalts from the Kerguelen Archipelago, southern Indian Ocean: Implications for Cenozoic eruption rates of the Kerguelen plume. *Earth and Planetary Science Letters* 174: 313–328.
Nielsen TK and Hopper JR (2004) From rift to drift: Mantle melting during continental breakup. *Geochemistry, Geophysics, Geosystems* 5.
Nielsen TK, Larsen HC, and Hopper JR (2002) Contrasting rifted margin styles south of Greenland; implications for mantle plume dynamics. *Earth and Planetary Science Letters* 200: 271–286.
Niu F, Solomon SC, Silver PG, Suetsugu D, and Inoue H (2002a) Mantle transition-zone structure beneath the South Pacific Superswell and evidence for a mantle plume underlying the Society Hotspot. *Earth and Planetary Science Letters* 198: 371–380.
Niu Y, Collerson KD, Batiza R, Wendt I, and Regelous M (1999) The origin of E-type MORB at ridges far from mantle plumes: The East Pacific Rise at 11°20′N. *Journal of Geophysical Research* 104: 7067–7087.
Niu Y, Regelous M, Wendt IJ, Batiza R, and O'Hara MJ (2002b) Geochemistry of near-EPR seamounts: Importance of source vs. process and the origin of enriched mantle component. *Earth and Planetary Science Letters* 199: 327–345.
Niu Y, Waggoner G, Sinton JM, and Mahoney JJ (1996) Mantle source heterogeneity and melting processes beneath seafloor spreading centers: The East Pacific Rise 18°-19°S. *Journal of Geophysical Research* 101: 27711–27733.
Nolet G, Allen R, and Zhao D (2007) Mantle plume tomography. *Chemical Geology* 241: 248–263.
Norton IO (2000) Global hotspot reference frames and plate motion. In: Richards MA, Gordon RG, and van der Hilst RD (eds.) *The History and Dynamics of Global Plate Motions*. *Geophysical Monograph*, vol. 121, pp. 339–357. Washington, DC: AGU.
Nyblade AA (2011) The upper-mantle low-velocity anomaly beneath Ethiopia, Kenya, and Tanzania: Constraints on the origin of the African superswell in eastern Africa and plate versus plume models of mantle dynamics. *Geological Society of America Special Papers* 478: 37–50.
Nyblade AA and Robinson SW (1994) The African superswell. *Geophysical Research Letters* 21: 765–768.
O'Connor JM and Duncan RA (1990) Evolution of the Walvis Ridge-Rio Grande Rise hot spot system: Implications for African and South American plate motions over plumes. *Journal of Geophysical Research* 95: 17475–17502.
O'Connor JM, Jokat W, le Roex AP, et al. (2012) Hotspot trails in the South Atlantic controlled by plume and plate tectonic processes. *Nature Geoscience* 5: 735–738.
O'Connor JM and Roex APL (1992) South Atlantic hot spot-plume systems: 1. Distribution of volcanism in time and space. *Earth and Planetary Science Letters* 113: 343–364.
O'Connor JM, Steinberger B, Regelous M, et al. (2013) Constraints on past plate and mantle motion from new ages for the Hawaiian-Emperor Seamount Chain. *Geochemistry, Geophysics, Geosystems* 14: 4564–4584.
O'Connor JM, Stoffers P, van den Bogaard P, and McWilliams M (1999) First seamount age evidence for significantly slower African Plate motion since 19 to 30 Ma. *Earth and Planetary Science Letters* 171: 575–589.
O'Connor JM, Stoffers P, and Wijbrans JR (2002) Pulsing of a focused mantle plume: Evidence from the distribution of Foundation Chain hotspot volcanism. *Geophysical Research Letters* 29(9): 4.
O'Connor JM, Stoffers P, and Wijbrans JR (2004) The Foundation Chain: Inferring hotspot-plate interaction from a weak seamount trail. In: Hekinian R, Stoffers P, and Cheminee J-L (eds.) *Oceanic Hotspots*, pp. 349–374. Springer.
O'Neill C, Müller D, and Steinberger B (2005) On the uncertainties in hot spot reconstructions and the significance of moving hot spot reference frames. *Geochemistry, Geophysics, Geosystems* 6: Q04003. http://dx.doi.org/10.1029/02004GC000784.
O'Reilly SY and Griffin WL (1988) Mantle Metasomatism beneath western Victoria, Australia.1. Metasomatic Processes in Cr-diopside Lherzolites. *Geochimica et Cosmochimica Acta* 52: 433–447.
Obrebski M, Allen RM, Xue M, and Hung S-H (2010) Slab-plume interaction beneath the Pacific Northwest. *Geophysical Research Letters* 37.
Ogden DE and Sleep NH (2012) Explosive eruption of coal and basalt and the end-Permian mass extinction. *Proceedings of the National Academy of Sciences of the United States of America* 109: 59–62.
Olson P (1990) Hot spots, swells and mantle plumes. In: Ryan MP (ed.) *Magma Transport and Storage*, pp. 33–51. New York: Wiley.
Olson P and Christensen U (1986) Solitary wave propagation in a fluid conduit within a viscous matrix. *Journal of Geophysical Research: Solid Earth* 91: 6367–6374.
Olson P, Schubert G, and Anderson C (1987) Plume formation in the D"-layer and the roughness of the core-mantle boundary. *Nature* 327: 409–413.
Olson P, Schubert G, and Anderson C (1993) Structure of axisymmetric mantle plumes. *Journal of Geophysical Research: Solid Earth* 98: 6829–6844.
Olson P, Schubert G, Anderson C, and Goldman P (1988) Plume formation and lithosphere erosion: A comparison of laboratory and numerical experiments. *Journal of Geophysical Research* 93: 15065–15084.
Olson P and Singer H (1985) Creeping plumes. *Journal of Fluid Mechanics* 158: 511–531.
Operto S and Charvis P (1996) Deep structure of the southern Kerguelen Plateau (southern Indian Ocean) from ocean bottom seismometer wide-angle seismic data. *Journal of Geophysical Research* 101: 25077–25103.
Oxburgh ER and Parmentier EM (1977) Compositional and density stratification in oceanic lithosphere-causes and consequences. *Geological Society of London* 133: 343–355.
Ozawa A, Tagami T, and Garcia MO (2005) Unspiked K-Ar dating of the Honolulu rejuvenated and Ko'olau shield volcanism on O'ahu, Hawai'i. *Earth and Planetary Science Letters* 232: 1–11.
Paczkowski K, Bercovici D, Landuyt W, and Brandon MT (2012) Drip instabilities of continental lithosphere: Acceleration and entrainment by damage. *Geophysical Journal International* 189: 717–729.

Pankhurst RJ, Leat PT, Sruoga P, et al. (1998) The Chon Aike province of Patagonia and related rocks in West Antarctica: A silicic large igneous province. *Journal of Volcanology and Geothermal Research* 81: 113–136.

Pankhurst RJ, Riley TR, Fanning CM, and Kelley SP (2000) Episodic silicic volcanism in Patagonia and the Antarctic Peninsula: Chronology of magmatism associated with the break-up of Gondwana. *Journal of Petrology* 41: 605–625.

Pares JM and Moore TC (2005) New evidence for the Hawaiian hotspot plume motion since the Eocene. *Earth and Planetary Science Letters* 237: 951–959.

Park Y and Nyblade AA (2006) P-wave tomography reveals a westward dipping low velocity zone beneath the Kenya Rift. *Geophysical Research Letters* 33.

Parkinson IJ, Shaefer BF, and Arculus RJ (2002) A lower mantle origin for the world's biggest LIP? A high precision Os isotope isochron from Ontong Java Plateau basalts drilled on ODP Leg 192. *Geochimica et Cosmochimica Acta* 66: A580.

Parman SW, Kurz MD, Hart SR, and Grove TL (2005) Helium solubility in olivine and implications for high $^{3}He/^{4}He$ in ocean island basalts. *Nature* 437(7062): 1140–1143.

Parnell-Turner RE, White NJ, Maclennan J, Henstock TJ, Murton BJ, and Jones SM (2013) Crustal manifestations of a hot transient pulse at 60°N beneath the Mid-Atlantic Ridge. *Earth and Planetary Science Letters* 363: 109–120.

Parsons B and Daly S (1983) The relationship between surface-topography, gravity-anomalies, and temperature structure of convection. *Journal of Geophysical Research* 88: 1129–1144.

Passier ML and Snieder RK (1996) Correlation between shear wave upper mantle structure and tectonic surface expressions: Application to central and southern Germany. *Journal of Geophysical Research* 101: 25293–25304.

Payne JL and Clapham ME (2012) End-permian mass extinction in the oceans: An ancient analog for the twenty-first century? In: Jeanloz R (ed.) *Annual Review of Earth and Planetary Sciences*, vol 40, pp. 89–111.

Peate DW (1997) The Parana-Etendeka provinces. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. *Geophysical Monograph*, 100, pp. 217–246. Washington, DC: American Geophysical Union.

Pertermann M and Hirschmann MM (2003) Partial melting experiments on a MORB-like pyroxenite between 2 and 3 GPa: Constraints on the presence of pyroxenite in basalt source regions from solidus locations and melting rate. *Journal of Geophysical Research* 108(B2). http://dx.doi.org/10.1029/2000JB 000118.

Peto MK, Mukhopadhyay S, and Kelley KA (2013) Heterogeneities from the first 100 million years recorded in deep mantle noble gases from the Northern Lau Back-arc Basin. *Earth and Planetary Science Letters* 369: 13–23.

Petterson MG (2004) The geology of north and central Malaita, Solomon Islands: The thickest and most accessible part of the world's largest (Ontong Java) oceanic plateau. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and evolution of the Ontong Java Plateau, Special publications 229*, pp. 63–82. London: Geological Society.

Phipps Morgan J (1999) Isotope topology of individual hotspot basalt arrays: Mixing curves or melt extraction trajectories. *Geochemistry, Geophysics, Geosystems* 1. http://dx.doi.org/10.1029/1999GC000004.

Phipps Morgan J (2001) Thermodynamics of pressure release melting of a veined plum pudding mantle. *Geochemistry, Geophysics, Geosystems* 2. http://dx.doi.org/10.1029/2000GC000049.

Phipps Morgan J, Morgan WJ, and Price E (1995a) Hotspot melting generates both hotspot volcanism and a hotspot swell. *Journal of Geophysical Research* 100: 8045–8062.

Phipps Morgan J, Morgan WJ, Zhang YS, and Smith WHF (1995b) Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection. *Journal of Geophysical Research* 100: 12753–12767.

Pietruszka AJ, Rubin KH, and Garcia MO (2001) (super 226) Ra- (super 230) Th- (super 238) U disequilibria of historical Kilauea lavas (1790–1982) and the dynamics of mantle melting within the Hawaiian Plume. *Earth and Planetary Science Letters* 186: 15–31.

Pik R, Deniel C, Coulon C, Yirgu G, Hofmann C, and Ayalew D (1998) The northwestern Ethiopian Plateau flood basalts. Classification and spatial distribution of magma types. *Journal of Volcanology and Geothermal Research* 81: 91–111.

Pilet S, Baker MB, and Stolper EM (2008) Metasomatized lithosphere and the origin of alkaline lavas. *Science* 320: 916–919.

Pilet S, Ulmer P, and Villiger S (2010) Liquid line of descent of a basanitic liquid at 1.5 Gpa: Constraints on the formation of metasomatic veins. *Contributions to Mineralogy and Petrology* 159: 621–643.

Pilidou S, Priestley K, Debayle E, and Gudmundsson O (2005) Rayleigh wave tomography in the North Atlantic: High resolution images of the Iceland, Azores and Eifel mantle plumes. *Lithos* 79: 453–474.

Pilidou S, Priestley K, Gudnmundsson O, and Debayle E (2004) Upper mantle S-wave speed heterogeneity and anisotropy beneath the North Atlantic from regional surface wave tomography: The Iceland and Azores plumes. *Geophysical Journal International* 159: 1057–1076.

Poore H, White N, and Maclennan J (2011) Ocean circulation and mantle melting controlled by radial flow of hot pulses in the Iceland plume. *Nature Geoscience* 4: 558–561.

Portnyagin M, Hoernle K, and Savelyev D (2009) Ultra-depleted melts from Kamchatkan ophiolites: Evidence for the interaction of the Hawaiian plume with an oceanic spreading center in the Cretaceous? *Earth and Planetary Science Letters* 287: 194–204.

Presnall DC, Gudfinnsson GH, and Walter MJ (2002) Generation of mid-ocean ridge basalts at pressures from 1 to 7 GPa. *Geochimica et Cosmochimica Acta* 66: 2073–2090.

Priestley K and Tilmann F (1999) Shear-wave structure of the lithosphere above the Hawaiian hot spot from two-station Rayleigh wave phase velocity measurements. *Geophysical Research Letters* 26: 1493–1496.

Putirka KD (2005) Mantle potential temperatures at Hawaii, Iceland, and the mid-ocean ridge system, as inferred from olivine phenocrysts: Evidence for thermally driven mantle plumes. *Geochemistry, Geophysics, Geosystems* 6: Q05L08. http://dx.doi.org/10.1029/2005GC000915.

Putirka KD, Perfit M, Ryerson FJ, and Jackson MG (2007) Ambient and excess mantle temperatures, olivine thermometry, and active vs. passive upwelling. *Chemical Geology* 241: 177–206.

Putirka KD, Ryerson FJ, Perfit M, and Ridley WI (2011) Mineralogy and composition of the oceanic mantle. *Journal of Petrology* 52: 279–313.

Raddick MJ, Parmentier EM, and Scheirer DS (2002) Buoyant decompression melting: A possible mechanisms for intra plate volcanism. *Journal of Geophysical Research* 107: 2228. http://dx.doi.org/10.1029/2001JB000617.

Ramalho R, Helffrich G, Cosca M, Vance D, Hoffmann D, and Schmidt DN (2010) Episodic swell growth inferred from variable uplift of the Cape Verde hotspot islands. *Nature Geoscience* 3: 774–777.

Ray JS, Mahoney JJ, Duncan RA, Ray J, Wessel P, and Naar DF (2012) Chronology and geochemistry of Lavas from the Nazca ridge and Easter Seamount chain: An similar to 30 Myr hotspot record. *Journal of Petrology* 53: 1417–1448.

Raymond C, Stock JM, and Cande SC (2000) Fast Paleogene motion of the Pacific hotspots from revised global plate circuit constraints. The history and dynamics of global plate motions. *AGU Geophysical Monograph* 121: 359–375.

Reichow MK, Pringle MS, Al'Mukhamedov AI, et al. (2009) The timing and extent of the eruption of the Siberian Traps large igneous province: Implications for the end-Permian environmental crisis. *Earth and Planetary Science Letters* 277: 9–20.

Reichow MK, Saunders AD, White RV, Al'Mukhamedov AL, and Medvedev AY (2005) Geochemistry and petrogenesis of basalts from the West Siberian basin: An extension of the Permo-Triassic Siberian Traps, Russia. *Lithos* 79: 425–452.

Reid MR, Bouchet RA, Blichert-Toft J, et al. (2012) Melting under the Colorado Plateau, USA. *Geology* 40: 387–390.

Reiners PW (2002) Temporal-compositional trends in intraplate basalt eruptions: Implications for mantle heterogeneity and melting processes. *Geochemistry, Geophysics, Geosystems* 3: 1011. http://dx.doi.org/10.1029/2002GC000250.

Renne PR, Deckart K, Ernesto M, Feraud G, and Piccirillo EM (1996) Age of the Ponta Grossa dike swarm (Brazil), and implications to Parana flood volcanism. *Earth and Planetary Science Letters* 144: 199–211.

Retallack GJ (2013) Permian and Triassic greenhouse crises. *Gondwana Research* 24: 90–103.

Reusch AM, Nyblade AA, Tibi R, et al. (2011) Mantle transition zone thickness beneath Cameroon: Evidence for an upper mantle origin for the Cameroon Volcanic Line. *Geophysical Journal International* 187: 1146–1150.

Reusch AM, Nyblade AA, Wiens DA, et al. (2010) Upper mantle structure beneath Cameroon from body wave tomography and the origin of the Cameroon Volcanic Line. *Geochemistry, Geophysics, Geosystems* 11.

Reyners M, Eberhart-Phillips D, and Bannister S (2011) Tracking repeated subduction of the Hikurangi Plateau beneath New Zealand. *Earth and Planetary Science Letters* 311: 165171.

Ribe N (1996) The dynamics of plume-ridge interaction 2. Off-ridge plumes. *Journal of Geophysical Research* 101: 16195–16204.

Ribe N, Christensen UR, and Theissing J (1995) The dynamics of plume-ridge interaction, 1: Ridge-centered plumes. *Earth and Planetary Science Letters* 134: 155–168.

Ribe N and Delattre WL (1998) The dynamics of plume-ridge interaction, 3: The effects of ridge migration. *Geophysical Journal International* 133: 511–518.

Ribe NM and Christensen UR (1994) Three-dimensional modelling of plume-lithosphere interaction. *Journal of Geophysical Research* 99: 669–682.

Ribe NM and Christensen UR (1999) The dynamical origin of Hawaiian volcanism. *Earth and Planetary Science Letters* 171: 517–531.

Ribe NM, Davaille A, and Christensen UR (2007) Fluid dynamics of mantle plumes. In: Ritter J and Christensen U (eds.) *Mantle Plumes – A Multidisciplinary Approach*, p. 148. New York: Springer.

Ribe NM and de Valpine DP (1994) The global hotspot distribution and instability of D" *Geophysical Research Letters* 21: 1507–1510.
Richard GC and Bercovici D (2009) Water-induced convection in the Earth's mantle transition zone. *Journal of Geophysical Research* 114.
Richards MA, Duncan RA, and Courtillot VE (1989) Flood basalts and hotspot tracks: Plume heads and tails. *Science* 246: 103–107.
Richards MA and Griffiths RW (1989) Thermal entrainment by deflected mantle plumes. *Nature* 342: 900–902.
Richter FM (1973) Convection and the large-scale circulation of the mantle. *Journal of Geophysical Research* 78: 8735–8745.
Richter FM and McKenzie DP (1981) On some consequences and possible causes of layered mantle convection. *Journal of Geophysical Research* 86: 6133–6142.
Rickers F, Fichtner A, and Trampert J (2013) The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion. *Earth and Planetary Science Letters* 367: 39–51.
Ritsema J and Allen RM (2003) The elusive mantle plume. *Earth and Planetary Science Letters* 207: 1–12.
Ritsema J, Deuss A, van Heijst HJ, and Woodhouse JH (2011) S40RTS: a degree-40 shear- velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophysical Journal International* 184: 1223–1236.
Ritsema J, van Heijst HJ, and Woodhouse JH (1999) Complex shear velocity structure beneath Africa and Island. *Science* 286: 1925–1928.
Ritsema J, van Heijst HJ, and Woodhouse JH (2004) Global transition zone tomography. *Journal of Geophysical Research* 109.
Ritter JRR, Jordan M, Christensen UR, and Achauer U (2001) A mantle plume below the Eifel volcanic fields, Germany. *Earth and Planetary Science Letters* 186: 7–14.
Ritzwoller MH, Shapiro NM, and Zhong S-J (2004) Cooling history of the Pacific lithosphere. *Earth and Planetary Science Letters* 226: 69–84.
Roberge J, Wallace PJ, White RV, and Coffin MF (2005) Anomalous uplift and subsidence of the Ontong Java Plateau inferred from CO_2 contents of submarine basaltic glasses. *Geology* 33: 501–504.
Roberge J, White RV, and Wallace PJ (2004) Volatiles in submarine basaltic glasses from the Ontong Java Plateau (ODP Leg 192): Implications for magmatic processes and source region compositions. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and Evolution of the Ontong Java Plateau*, pp. 239–258. London: Geological Society, Special Publications 229.
Rogers GC (1982) Oceanic plateaus as meteorite impact signatures. *Nature* 299: 341–342.
Rooney TO, Herzberg C, and Bastow ID (2012) Elevated mantle temperature beneath East Africa. *Geology* 40: 27–30.
Roy AB (2003) Geological and geophysical manifestations of the Reunion Plume-Indian lithosphere interactions: Evidence from northwest India. *Gondwana Research* 6: 487–500.
Rychert CA, Laske G, Harmon N, and Shearer PM (2013) Seismic imaging of melt in a displaced Hawaiian plume. *Nature Geoscience* 6: 657–660.
Saal AE, Hart SR, Shimizu N, Hauri EH, and Layne GD (1998) Pb isotopic variability in melt inclusions from oceanic island basalt, Polynesia. *Science* 282: 1481–1484.
Saffman PG and Taylor GI (1958) The penetration of a fluid into a porous medium of Hele-Shaw cell containing a more viscous liquid. *Proceedings of the Royal Society of London* A245: 312–329.
Sager WW (2005) What built Shatsky Rise, a mantle plume or ridge tectonics? In: Foulger G, Natland JH, Presnall DC, and Anderson DL (eds.) *Plumes, Plates, and Paradigms Special Paper 388*. Boulder, CO: Geological Society of America.
Sager WW, Kim J, Klaus A, Nakanishi M, and Khankishieva LM (1999) Bathymetry of Shatsky Rise, northwest Pacific Ocean: Implications for ocean plateau development at a triple junction. *Journal of Geophysical Research: Solid Earth* 104: 7557–7576.
Sager WW, Lamarch AJ, and Kopp C (2005) Paleomagnetic modeling of seamounts near the Hawaiian-Emperor bend. *Tectonophysics* 405: 121–140.
Sager WW, Zhang J, Korenaga J, et al. (2013) An immense shield volcano within the Shatsky Rise oceanic plateau, northwest Pacific Ocean. *Nature Geoscience* 6: 976–981, advance online publication.
Sakamaki T, Suzuki A, Ohtani E, et al. (2013) Ponded melt at the boundary between the lithosphere and asthenosphere. *Nature Geoscience* 6: 1041–1044.
Sallares V, Charvis P, Flueh ER, and Bialas J (2005) Seismic structure of the Carnegie ridge and the nature of the Galapagos hotspot. *Geophysical Journal International* 161: 763–788.
Salters VJM and Dick HJB (2002) Mineralogy of the mid-ocean-ridge basalt source from neodymium and isotopic composition of abyssal peridotites. *Nature* 418: 68–72.
Saltzer RL and Humphreys ED (1997) Upper mantle P wave velocity structure of the eastern Snake River Plain and its relationship to geodynamic models of the region. *Journal of Geophysical Research* 102: 11829–11841.
Samuel H and Bercovici D (2006) Oscillating and stagnating plumes in the Earth's lower mantle. *Earth and Planetary Science Letters* 248: 90–105.
Sandwell D and Fialko Y (2004) Warping and cracking of the Pacific plate by thermal contraction. *Journal of Geophysical Research* 109: B10411. http://dx.doi.org/10.1029/12004JB0003091.
Sandwell DT, Winterer EL, Mammerickx J, et al. (1995) Evidence for diffuse extension of the Pacific plate from Pukapuka ridges and cross-grain gravity lineations. *Journal of Geophysical Research* 100: 15087–15099.
Saunders AD, England RW, Relchow MK, and White RV (2005) A mantle plume origin for the Siberian traps: Uplift and extension in the West Siberian Basin, Russia. *Lithos* 79: 407–424.
Saunders AD, Fitton JG, Kerr AC, Norry MJ, and Kent RW (1997) The North Atlantic igneous province. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. *Geophysical Monograph*, vol. 100, pp. 45–94. Washington, DC: American Geophysical Union.
Saunders AD, Jones SM, Morgan LA, Pierce KL, Widdowson M, and Xu YG (2007) Regional uplift associated with continental large igneous provinces: The roles of mantle plumes and the lithosphere. *Chemical Geology* 241: 282–318.
Schaller MF, Wright JD, and Kent DV (2011) Atmospheric P-CO_2 perturbations associated with the central atlantic magmatic province. *Science* 331: 1404–1409.
Schaller MF, Wright JD, Kent DV, and Olsen PE (2012) Rapid emplacement of the Central Atlantic Magmatic Province as a net sink for CO_2. *Earth and Planetary Science Letters* 323: 27–39.
Schaltegger U, Guex J, Bartolini A, Schoene B, and Ovtcharova M (2008) Precise U-Pb age constraints for end-Triassic mass extinction, its correlation to volcanism and Hettangian post-extinction recovery. *Earth and Planetary Science Letters* 267: 266–275.
Schilling J-G (1971) Sea floor evolution: Rare earth evidence. *Philosophical Transactions of the Royal Society of London A* 268A: 663–706.
Schilling J-G (1973) Iceland mantle plume: Geochemical study of Reykjanes Ridge. *Nature* 242: 565–571.
Schilling J-G (1991) Fluxes and excess temperatures of mantle plumes inferred from their interaction with migrating mid-ocean ridges. *Nature* 352: 397–403.
Schilling J-G, Thompson G, Kingsley R, and Humphris S (1985) Hotspot-migrating ridge interaction in the South Atlantic. *Nature* 313: 187–191.
Schmandt B, Dueker K, Humphreys E, and Hansen S (2012) Hot mantle upwelling across the 660 beneath Yellowstone. *Earth and Planetary Science Letters* 331: 224–236.
Schmandt B and Humphreys E (2010) Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth and Planetary Science Letters* 297: 435–445.
Schmerr N (2012) The gutenberg discontinuity: Melt at the lithosphere-asthenosphere boundary. *Science* 335: 1480–1483.
Schmerr N, Garnero E, and McNamara A (2010) Deep mantle plumes and convective upwelling beneath the Pacific Ocean. *Earth and Planetary Science Letters* 294: 143–151.
Schubert G, Masters G, Olson P, and Tackley P (2004) Superplumes or plume clusters? *Physics of the Earth and Planetary Interiors* 146: 147–162.
Schubert G and Olson P (1989) Solitary waves in mantle plumes. *Journal of Geophysical Research* 94: 9523–9532.
Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*, p. 940. Cambridge: Cambridge University Press.
Schulte P, Alegret L, Arenillas I, et al. (2010) The chicxulub asteroid impact and mass extinction at the cretaceous-paleogene boundary. *Science* 327: 1214–1218.
Schutt DL and Lesher CE (2006) Effects of melt depletion on the density and seismic velocity of garnet and spinel lherzolite. *Journal of Geophysical Research* 111.
Scott DR and Stevenson DJ (1989) A self-consistent model for melting, magma migration and buoyancy-driven circulation beneath mid-ocean ridges. *Journal of Geophysical Research* 94: 2973–2988.
Scott DR, Stevenson DJ, and Whitehead JA (1986) Observations of solitary waves in a viscously deformable pipe. *Nature* 319: 759–761.
Self S, Widdowson M, Thordarson T, and Jay AE (2006) Volatile fluxes during flood basalt eruptions and potential effects on the global environment: A Deccan perspective. *Earth and Planetary Science Letters* 248: 518–532.
Sharma M (1997) Siberian traps. In: Mahoney JJ and Coffi MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism*. *Geophysical Monograph*, 100, pp. 273–296. Washington, DC: American Geophysical Union.
Sharp WD and Clague DA (2006) 50-Ma initiation of Hawaiian-Emperor bend records major change in Pacific plate motion. *Science* 313: 1281–1284.
Shellnutt JG, Denyszyn SW, and Mundil R (2012) Precise age determination of mafic and felsic intrusive rocks from the Permian Emeishan large igneous province (SW China). *Gondwana Research* 22: 118–126.

Shellnutt JG and Jahn BM (2011) Origin of Late Permian Emeishan basaltic rocks from the Panxi region (SW China): Implications for the Ti-classification and spatial-compositional distribution of the Emeishan flood basalts. *Journal of Volcanology and Geothermal Research* 199: 85–95.

Shen Y and Forsyth DW (1995) Geochemical constraints on initial and final depths of melting beneath mid-ocean ridges. *Journal of Geophysical Research* 100: 2211–2237.

Shen Y, Solomon SC, Bjarnason IT, et al. (2002) Seismic evidence for a tilted mantle plume and north-south mantle flow beneath Iceland. *Earth and Planetary Science Letters* 197: 261–272.

Shen Y, Solomon SC, Bjarnason IT, and Wolfe C (1998) Seismic evidence for a lower mantle origin of the Iceland plume. *Nature* 395: 62–65.

Shervais JW and Hanan BB (2008) Lithospheric topography, tilted plumes, and the track of the Snake River-Yellowstone hot spot. *Tectonics* 27.

Sheth HC (2005) from deccan to reunion: No trace of a mantle plume. In: Foulger G, Natland JH, Presnall DC, and Anderson DL (eds.) *Plumes, Plates, and Paradigms Special Paper 388*, pp. 477–501. Geological Society of America.

Shorttle O and Maclennan J (2011) Compositional trends of Icelandic basalts: Implications for short-length scale lithological heterogeneity in mantle plumes. *Geochemistry, Geophysics, Geosystems* 12.

Shorttle O, Maclennan J, and Jones SM (2010) Control of the symmetry of plume-ridge interaction by spreading ridge geometry. *Geochemistry, Geophysics, Geosystems* 11.

Sigloch K (2011) Mantle provinces under North America from multifrequency P wave tomography. *Geochemistry, Geophysics, Geosystems* 12.

Sikora PJ and Bergen JA (2004) Lower Cretaceous planktonic foraminiferal and nannofossil biostratigraphy of Ontong Java Plateau sites from DSDP Leg 30 and ODP Leg 192. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and evolution of the Ontong Java Plateau, Special Publications 229*, pp. 83–112. London: Geological Society.

Silva IGN, Weis D, Scoates JS, and Barling J (2013) The ninetyeast ridge and its relation to the kerguelen, Amsterdam and St. Paul hotspots in the Indian ocean. *Journal of Petrology* 54: 1177–1210.

Simmons NA, Forte AM, and Grand SP (2006) Constraining mantle flow with seismic and geodynamic data: A joint approach. *Earth and Planetary Science Letters* 246: 109–124.

Sims KWW, DePaolo DJ, Murrell MT, et al. (1999) Porosity of the melting zone and variations in the solid mantle upwelling rate beneath Hawaii: Inferences from U-238-Th-230-Ra-226 and U-235-Pa-231 disequilibria. *Geochimica et Cosmochimica Acta* 63: 4119–4138.

Singh SC, Carton H, Chauhan AS, et al. (2011) Extremely thin crust in the Indian Ocean possibly resulting from Plume-Ridge Interaction. *Geophysical Journal International* 184: 29–42.

Sinton CW, Christie DM, and Duncan RA (1996) Geochronology of the Galapagos seamounts. *Journal of Geophysical Research* 101: 13689–13700.

Sinton CW, Duncan RA, and Denyer P (1997) Nicoya Peninsula, Costa Rica: A single suite of Caribbean oceanic plateau magmas. *Journal of Geophysical Research* 102: 15507–15520.

Skilbeck JN and Whitehead JA (1978) Formation of discrete islands in linear island chains. *Nature* 272: 499–501.

Slater L, Jull M, McKenzie D, and Gronvold K (1998) Deglaciation effects on mantle melting under Iceland: Results from the northern volcanic zone. *Earth and Planetary Science Letters* 164: 151–164.

Sleep NH (1990) Hotspots and mantle plumes: Some phenomenology. *Journal of Geophysical Research* 95: 6715–6736.

Sleep NH (1996) Lateral flow of hot plume material ponded at sublithospheric depths. *Journal of Geophysical Research* 101: 28065–28083.

Sleep NH (2006) Mantle plumes from top to bottom. *Earth-Science Reviews* 77: 231–271.

Sleep NH (2007) Edge-modulated stagnant-lid convection and volcanic passive margins. *Geochemistry, Geophysics, Geosystems* 8.

Sleep NH (2008) Channeling at the base of the lithosphere during the lateral flow of plume material beneath flow line hot spots. *Geochemistry, Geophysics, Geosystems* 9.

Sleep NH (2011) Seismically observable features of mature stagnant-lid convection at the base of the lithosphere: Some scaling relationships. *Geochemistry, Geophysics, Geosystems* 12.

Smallwood JR and White RS (2002) Ridge-plume interaction in the North Atlantic and its influence on continental breakup and seafloor spreading. *Geological Society* 197: 15–37, Special Publications.

Smirnov AV and Tarduno JA (2010) Co-location of eruption sites of the Siberian Traps and North Atlantic Igneous Province: Implications for the nature of hotspots and mantle plumes. *Earth and Planetary Science Letters* 297: 687–690.

Smith RB and Braile LW (1994) The Yellowstone hotspot. *Journal of Volcanology and Geothermal Research* 61: 121–187.

Smith WHF and Sandwell DT (1997) Global seafloor topography from satellite altimetry and ship depth soundings. *Science* 277: 1957–1962.

Sobolev AV, Hofmann AW, Bruegmann G, Batanova VG, and Kuzmin DV (2008) A quantitative link between recycling and osmium isotopes. *Science* 321: 536.

Sobolev AV, Hofmann AW, Kuzmin DV, et al. (2007) The amount of recycled crust in sources of mantle-derived melts. *Science* 316: 412–417.

Sobolev AV, Hofmann AW, Sobolev SV, and Nikogosian IK (2005) An olivine-free mantle source of Hawaiian shield basalts. *Nature* 434: 590–597.

Sobolev AV, Krivolutskaya NA, and Kuzmin DV (2009) Petrology of the parental melts and mantle sources of Siberian trap magmatism. *Petrology* 17: 253–286.

Sobolev SV, Sobolev AV, Kuzmin DV, et al. (2011) Linking mantle plumes, large igneous provinces and environmental catastrophes. *Nature* 477: U312–U380.

Spandler C, Yaxley G, Green DH, and Rosenthal A (2008) Phase relations and melting of anhydrous k-bearing eclogite from 1200 to 1600 °C and 3 to 5 GPa. *Journal of Petrology* 49: 771–795.

Sparks DW, Parmentier EM, and Morgan JP (1993) Three-dimensional mantle convection beneath a segmented spreading center: Implications for along-axis variations in crustal thickness and gravity. *Journal of Geophysical Research* 98: 21977–21995.

Staudigel H, Park KH, Pringle MS, Rubenstone JL, Smith WHF, and Zindler A (1991) The longevity of the South Pacific isotopic and thermal anomaly. *Earth and Planetary Science Letters* 102: 24–44.

Stein CA and Stein S (1992) A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* 359: 123–129.

Stein CA and Stein S (1993) Constraints on Pacific midplate swells from global depth-age and heat flow-age models. In: Pringle M, Sager W, Sliter W, and Stein S (eds.) *The Mesozoic Pacific. Geophysical Monograph*, vol. 76, pp. 53–76. American Geophysical Union.

Stein CA and Stein S (1994) Constraints on hydrothermal heat flux through the oceanic lithosphere from global heat flow. *Journal of Geophysical Research* 99: 3081–3095.

Stein CA and Stein S (2003) Mantle plumes: Heat-flow near Iceland. *Astronomy and Geophysics* 44: 8–10.

Steinberger B (2000) Plumes in a convecting mantle: Models and observations for individual hotspots. *Journal of Geophysical Research* 105: 11127–11152.

Steinberger B and Antretter M (2006) Conduit diameter and buoyant rising speed of mantle plumes: Implications for the motion of hot spots and shape of plume conduits. *Geochemistry, Geophysics, Geosystems* 7.

Steinberger B, Sutherland R, and O'Connell RJ (2004) Prediction of Emperor-Hawaii seamount locations from a revised model of global plate motion and mantle flow. *Nature* 430: 167–173.

Steinberger B and Torsvik TH (2012) A geodynamic model of plumes from the margins of Large Low Shear Velocity Provinces. *Geochemistry, Geophysics, Geosystems* 13.

Stewart K and Rogers N (1996) Mantle plume and lithosphere contributions to basalts from southern Ethiopia. *Earth and Planetary Science Letters* 139: 195–211.

Stewart K, Turner S, Kelley S, Hawkesworth C, Kirstein L, and Mantovani M (1996) 3-D, (super 40) Ar- (super 39) Ar geochronology in the Parana continental flood basalt province. *Earth and Planetary Science Letters* 143: 95–109.

Stixrude L and Lithgow-Bertelloni C (2005) Mineralogy and elasticity of the oceanic upper mantle: Origin of the low-velocity zone. *Journal of Geophysical Research* 110.

Storey M, Duncan RA, and Tegner C (2007) Timing and duration of volcanism in the North Atlantic Igneous Province: Implications for geodynamics and links to the Iceland hotspot. *Chemical Geology* 241: 264–281.

Storey M, Mahoney JJ, and Saunders AD (1997) Cretaceous basalts in Madagascar and the transition between plume and continental lithosphere mantle sources. In: Mahoney JJ and Coffin MF (eds.) *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism. Geophysical Monograph*, 100, pp. 95–122. Washington, DC: American Geophysical Union.

Stracke A, Zindler A, Salters VJM, et al. (2003) Theistareykir revisited. *Geochemistry, Geophysics, Geosystems* 4: 8507. http://dx.doi.org/10.1029/2001GC000201.

Stuart FM, Lass-Evans S, Fitton JG, and Ellam RM (2003) High $^{3}He/^{4}He$ ratios in pictritic basalts from Baffin Island and the role of a mixed reservoir in mantle plumes. *Nature* 424: 57–59.

Styles E, Goes S, van Keken PE, Ritsema J, and Smith H (2011) Synthetic images of dynamically predicted plumes and comparison with a global tomographic model. *Earth and Planetary Science Letters* 311: 351–363.

Suetsugu D, Isse T, Tanaka S, et al. (2009) South Pacific mantle plumes imaged by seismic observation on islands and seafloor. *Geochemistry, Geophysics, Geosystems* 10.

Sun Y, Lai X, Wignall PB, et al. (2010) Dating the onset and nature of the Middle Permian Emeishan large igneous province eruptions in SW China using conodont biostratigraphy and its bearing on mantle plume uplift models. *Lithos* 119: 20–33.
Svensen H, Corfu F, Polteau S, Hammer O, and Planke S (2012) Rapid magma emplacement in the Karoo Large Igneous Province. *Earth and Planetary Science Letters* 325: 1–9.
Svensen H, Planke S, Polozov AG, et al. (2009) Siberian gas venting and the end-Permian environmental crisis. *Earth and Planetary Science Letters* 277: 490–500.
Tackley PJ (2011) Living dead slabs in 3-D: The dynamics of compositionally-stratified slabs entering a "slab graveyard" above the core-mantle boundary. *Physics of the Earth and Planetary Interiors* 188: 150–162.
Tackley PJ and Stevenson DJ (1993) A mechanism for spontaneous self-perpetuating volcanism on terrestrial planets. In: Stone DB and Runcorn SK (eds.) *Flow and Creep in the Solar System: Observations, Modeling and Theory*, pp. 307–321. Norwell, MA: Kluwer Academic.
Takahashi E (2002) The Hawaiian plume and magma genesis. *Geophysical Monograph* 128: 347–348.
Takahashi E and Kushiro I (1983) Melting of a dry peridotite at high-pressures and basalt magma genesis. *American Mineralogist* 68: 859–879.
Takahashi E, Nakajima K, and Wright TL (1998) Origin of the Columbia Rivers basalts: Melting model of a heterogeneous plume head. *Earth and Planetary Science Letters* 162: 63–80.
Takazawa E, Frey FA, Shimizu N, and Obata M (2000) Whole rock compositional variations in an upper mantle peridotite (Horoman, Hokkaido, Japan): Are they consistent with a partial melting process? *Geochimica et Cosmochimica Acta* 64: 695–716.
Tan E and Gurnis M (2007) Compressible thermochemical convection and application to lower mantle structures. *Journal of Geophysical Research* 112.
Tan E, Leng W, Zhong S, and Gurnis M (2011) On the location of plumes and lateral movement of thermochemical structures with high bulk modulus in the 3-D compressible mantle. *Geochemistry, Geophysics, Geosystems* 12.
Tanaka S, Obayashi M, Suetsugu D, et al. (2009) P-wave tomography of the mantle beneath the South Pacific Superswell revealed by joint ocean floor and islands broadband seismic experiments. *Physics of the Earth and Planetary Interiors* 172: 268–277.
Tarduno J, Bunge H-P, Sleep N, and Hansen U (2009) The Bent Hawaiian-Emperor hotspot track: Inheriting the mantle wind. *Science* 324: 50–53.
Tarduno JA, Duncan RA, Scholl DW, et al. (2003) The Emperor Seamounts; southward motion of the Hawaiian Hotspot plume in Earth's mantle. *Science* 301: 1064–1069.
Tarduno JA and Gee J (1995) Large-scale motion between Pacific and Atlantic hotspots. *Nature* 378: 477–480.
Tarduno JA, Sliter WV, Kroenke LW, et al. (1991) Rapid formation of Ontong Java Plateau by Aptian mantle plume volcanism. *Science* 254: 399–403.
Tauzin B, Debayle E, and Wittlinger G (2008) The mantle transition zone as seen by global Pds phases: No clear evidence for a thin transition zone beneath hotspots. *Journal of Geophysical Research* 113.
Taylor B (2006) The single largest oceanic plateau: Ontong Java-Manihiki-Hikurangi. *Earth and Planetary Science Letters* 241: 372–380.
Tejada MLG, Mahoney JJ, Castillo PR, Ingle SP, Sheth HC, and Weis D (2004) Pin-pricking the elephant: Evidence on the origin of the Ontong Java Plateau from Pb-Sr-Hf-Nd isotopic characteristics of ODP Leg 192. In: Fitton JG, Mahoney JJ, Wallace PJ, and Saunders AD (eds.) *Origin and Evolution of the Ontong Java Plateau, Special Publications 229*, pp. 133–150. London: Geological Society.
Tejada MLG, Mahoney JJ, Duncan RA, and Hawkins MP (1996) Age and geochemistry of basement and alkalic rocks of Malaita and Santa Isabel, Solomon Islands, southern margin of Ontong Java Plateau. *Journal of Petrology* 37: 361–394.
Tejada MLG, Mahoney JJ, Neal CR, Duncan RA, and Petterson MG (2002) Basement geochemistry and geochronology of central Malaita and Santa Isabel, Solomon Islands, southern margin of Ontong Java Plateau. *Journal of Petrology* 37: 449–484.
ten Brink U (1991) Volcano spacing and plate rigidity. *Geology* 19: 397–400.
Thompson RN and Gibson SA (1991) Subcontinental mantle plumes, hotspots and preexisting thinspots. *Journal of the Geological Society* 148: 973–977.
Thorne MS, Garnero EJ, and Grand SP (2004) Geographic correlation between hot spots and deep mantle lateral shear-wave velocity gradients. *Physics of the Earth and Planetary Interiors* 146: 4763.
Tian Y and Zhao D (2012) P-wave tomography of the western United States: Insight into the Yellowstone hotspot and the Juan de Fuca slab. *Physics of the Earth and Planetary Interiors* 200: 72–84.
Till CB, Elkins-Tanton LT, and Fischer KM (2010) A mechanism for low-extent melts at the lithosphere-asthenosphere boundary. *Geochemistry, Geophysics, Geosystems* 11.
Tilmann FJ, Benz HM, Priestley KF, and Okubo PG (2001) P-wave velocity structure of the uppermost mantle beneath Hawaii from traveltime tomography. *Geophysical Journal International* 146: 594–606.
Tiwari VM, Grevemeyer I, Singh B, and Morgan JP (2007) Variation of effective elastic thickness and melt production along the Deccan-Reunion hotspot track. *Earth and Planetary Science Letters* 264: 9–21.
To A, Romanowicz B, Capdeville Y, and Takeuchi N (2005) 3D effects of sharp boundaries at the borders of the African and Pacific Superplumes: Observation and modeling. *Earth and Planetary Science Letters* 233: 137–153.
Toomey DR, Hooft EEE, and Detrick R (2001) Crustal thickness variations and internal structure of the Galapagos Archipelago. *EOS, Transactions of the American Geophysical Union* 82(47), Fall Meeting Supplement, Abstract T42B-0939.
Torsvik TH, Burke K, Steinberger B, Webb SJ, and Ashwal LD (2010a) Diamonds sampled by plumes from the core-mantle boundary. *Nature* 466: U100–U352.
Torsvik TH, Smethurst MA, Burke K, and Steinberger B (2006) Large igneous provinces generated from the margins of the large low-velocity provinces in the deep mantle. *Geophysical Journal International* 167: 1447–1460.
Torsvik TH, Steinberger B, Gurnis M, and Gaina C (2010b) Plate tectonics and net lithosphere rotation over the past 150 My. *Earth and Planetary Science Letters* 291: 106–112.
Torsvik TH, Van der Voo R, Preeden U, et al. (2012) Phanerozoic polar wander, palaeogeography and dynamics. *Earth-Science Reviews* 114: 325–368.
Torsvik TH, Van der Voo R, and Redfield TF (2002) Relative hotspot motions versus true polar wander. *Earth and Planetary Science Letters* 202: 185–200.
Tosi N and Yuen DA (2011) Bent-shaped plumes and horizontal channel flow beneath the 660 km discontinuity. *Earth and Planetary Science Letters* 312: 348–359.
Trampert J, Deschamps F, Resovsky J, and Yuen D (2004) Probabilistic tomography maps chemical heterogeneities throughout the lower mantle. *Science* 306: 853–856.
Trumbull RB, Vietor T, Hahne K, Wackerle R, and Ledru P (2004) Aeromagnetic mapping and reconnaissance geochemistry of the Early Cretaceous Henties Bay-Outjo dike swarm, Etendeka Igneous Province, Namibia. *Journal of African Earth Sciences* 40: 17–29.
Turner DL, Jarrard RD, and Forbes RB (1980) Geochronology and origin of the Pratt-Welker seamount chain, Gulf of Alaska: A new pole of rotation for the Pacific plate. *Journal of Geophysical Research* 85: 6547–6556.
Turner S, et al. (1994) Magmatism and continental bread-up in the South Atlantic: High precision ^{40}Ar-^{39}Ar geochronology. *Earth and Planetary Science Letters* 121: 333–348.
Uenzelmann-Neben G, Schmidt DN, Niessen F, and Stein R (2012) Intraplate volcanism off South Greenland: Caused by glacial rebound? *Geophysical Journal International* 190: 1–7.
van Ark E and Lin J (2004) Time variation in igneous volume flux of the Hawaii-Emperor seamount chain. *Journal of Geophysical Research* 109: B11401. http://dx.doi.org/10.1029/2003JB002949.
van der Hilst RD and Karason H (1999) Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model. *Science* 283: 1885–1888.
van Hinsbergen DJJ, Steinberger B, Doubrovine PV, and Gassmoeller R (2011) Acceleration and deceleration of India-Asia convergence since the Cretaceous: Roles of mantle plumes and continental collision. *Journal of Geophysical Research* 116.
van Hunen J and Zhong S (2003) New insight in the Hawaiian plume swell dynamics from scaling laws. *Geophysical Research Letters* 30: 1785. http://dx.doi.org/10.1029/2003GL017646.
van Hunen J, Zhong SJ, Shapiro NM, and Ritzwoller MH (2005) New evidence for dislocation creep from 3-D geodynamic modeling of the Pacific upper mantle structure. *Earth and Planetary Science Letters* 238: 146–155.
van Keken PE (1997) Evolution of starting mantle plumes: A comparison between numerical and laboratory models. *Earth and Planetary Science Letters* 148: 1–11.
van Keken PE, Davaille A, and Vatteville J (2013) Dynamics of a laminar plume in a cavity: The influence of boundaries on the steady state stem structure. *Geochemistry, Geophysics, Geosystems* 14: 158–178.
van Keken PE and Gable CW (1995) The interaction of a plume with a rheological boundary: A comparison between two- and three-dimensional models. *Journal of Geophysical Research* 100: 20291–20302.
van Wijk JW, Baldridge WS, van Hunen J, et al. (2010) Small-scale convection at the edge of the Colorado Plateau: Implications for topography, magmatism, and evolution of Proterozoic lithosphere. *Geology* 38: 611–614.
van Wijk JW, Huismans RS, ter Voorde M, and Cloetingh S (2001) Melt generation at volcanic continental margins: No need for a mantle plume? *Geophysical Research Letters* 28: 3995–3998.

Vandamme D, Courtillot V, Besse J, and Montinny R (1991) Paleomagnetism and age determinations of the Deccan Traps (India): Results of a Nagpur-Bombay traverse and review of earlier work. *Reviews of Geophysics* 29: 159–190.

VanDecar JC, James DE, and Assumpcao M (1995) Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces. *Nature* 378: 25–31.

Vatteville J, van Keken PE, Limare A, and Davaille A (2009) Starting laminar plumes: Comparison of laboratory and numerical modeling. *Geochemistry, Geophysics, Geosystems* 10.

Vdovin O, Rial JA, Levshin AL, and Ritzwoller MH (1999) Group-velocity tomography of South America and the surrounding oceans. *Geophysical Journal International* 136: 324–340.

Verati C, Rapaille C, Feraud G, Marzoli A, Bertrand H, and Youbi N (2007) Ar-40/Ar-39 ages and duration of the Central Atlantic Magmatic Province volcanism in Morocco and Portugal and its relation to the Triassic-Jurassic boundary. *Palaeogeography, Palaeoclimatology, Palaeoecology* 244: 308–325.

Vidal V and Bonneville A (2004) Variations of the Hawaiian hot spot activity revealed by variations in magma production rate. *Journal of Geophysical Research* 109. http://dx.doi.org/10.1029/2003JB002559.

Villagomez DR, Toomey DR, Hooft EEE, and Solomon SC (2007) Upper mantle structure beneath the Galapagos Archipelago from surface wave tomography. *Journal of Geophysical Research* 112.

Vinnik L, Silveira G, Kiselev S, Farra V, Weber M, and Stutzmann E (2012) Cape Verde hotspot from the upper crust to the top of the lower mantle. *Earth and Planetary Science Letters* 319: 259–268.

Vlastelic I and Dosso L (2005) Initiation of a plume-ridge interaction in the South Pacific recorded by high-precision Pb isotopes along Hollister Ridge. *Geochemistry, Geophysics, Geosystems* 6.

Waite GP, Smith RB, and Allen RM (2006) V-P and V-S structure of the Yellowstone hot spot from teleseismic tomography: Evidence for an upper mantle plume. *Journal of Geophysical Research* 111.

Walker KT, Bokelmann GHR, Klemperer SL, and Bock G (2005) Shear-wave splitting around the Eifel hotspot: Evidence for a mantle upwelling. *Geophysical Journal International* 163: 962–980.

Wang S and Wang R (2001) Current plate velocities relative to hotspots: Implications for hotspot motion, mantle viscosity and global reference frame. *Earth and Planetary Science Letters* 189: 133–140.

Watson S and McKenzie D (1991) Melt generation by plumes – A study of Hawaiian volcanism. *Journal of Petrology* 32: 501–537.

Watts AB, Weissel JK, Duncan RA, and Larson RL (1988) Origin of the Louisville ridge and its relationship to the Eltanin fracture zone system. *Journal of Geophysical Research* 93: 3051–3077.

Weeraratne DS, Forsyth DW, Yang Y, and Webb SC (2007) Rayleigh wave tomography beneath intraplate volcanic ridges in the South Pacific. *Journal of Geophysical Research* 112.

Weis D, Frey FA, Schlich R, et al. (2002) Trace of the Kerguelen mantle plume: Evidence from seamounts between the Kerguelen Archipelago and Heard Island, Indian Ocean. *Geochemistry, Geophysics, Geosystems* 3.

Weis D, Garcia MO, Rhodes JM, Jellinek M, and Scoates JS (2011) Role of the deep mantle in generating the compositional asymmetry of the Hawaiian mantle plume. *Nature Geoscience* 4: 831–838.

Wen L (2006) A compositional anomaly at the Earth's core-mantle boundary as an anchor to the relatively slowly moving surface hotspots and as source to the DUPAL anomaly. *Earth and Planetary Science Letters* 246: 138–148.

Werner R, Hoernle K, van den Bogaard P, Ranero C, von Huene R, and Korich D (1999) Drowned 14-m.y.-old Galapagos Archipelago off the coast of Costa Rica: Implications for tectonic and evolutionary models. *Geology* 27: 499–502.

Wessel P (1993) Observational constraints on models of the Hawaiian hot spot swell. *Journal of Geophysical Research* 98: 16095–16104.

Wessel P and Kroenke LW (2008) Pacific absolute plate motion since 145 Ma: An assessment of the fixed hot spot hypothesis. *Journal of Geophysical Research* 113.

Wessel P and Kroenke LW (2009) Observations of geometry and ages constrain relative motion of Hawaii and Louisville plumes. *Earth and Planetary Science Letters* 284: 467–472.

Wessel P and Smith WHF (1995) New version of the Generic Mapping Tools released. *EOS, Transactions of the American Geophysical Union* 76: 329.

West JD, Fouch MJ, Roth JB, and Elkins-Tanton LT (2009) Vertical mantle flow associated with a lithospheric drip beneath the Great Basin. *Nature Geoscience* 2: 438–443.

White R and McKenzie D (1989) Magmatism at rift zones: The generation of volcanic continental margins and flood basalts. *Journal of Geophysical Research* 94: 7685–7729.

White RS (1988) A hot-spot model for early Tertiary volcanism in the N Atlantic. *Geological Society* 39: 3–13, special publications.

White RS (1997) Rift-plume interaction in the North Atlantic. *Philosophical Transactions of the Royal Society of London, Series A* 355: 319–339.

White RS, Bown JW, and Smallwood JR (1995) The temperature of the Iceland plume and origin of outward propagating V-shaped ridges. *Journal of the Geological Society* 152: 1039–1045.

White RS and McKenzie D (1995) Mantle plumes and flood basalts. *Journal of Geophysical Research* 100: 17543–17585.

Whitehead JA and Helfrich KR (1988) Wave transport of deep mantle material. *Nature* 336: 59–61.

Whitehead JA and Luther DS (1975) Dynamics of laboratory diapir and plume models. *Journal of Geophysical Research* 80: 705–717.

Whiteside JH, Olsen PE, Kent DV, Fowell SJ, and Et-Touhami M (2007) Synchrony between the Central Atlantic magmatic province and the Triassic-Jurassic mass-extinction event? *Palaeogeography, Palaeoclimatology, Palaeoecology* 244: 345–367.

Whittaker JM, Mueller RD, Leitchenkov G, et al. (2007) Major Australian-Antarctic plate reorganization at Hawaiian-Emperor bend time. *Science* 318: 83–86.

Williams Q, Revenaugh J, and Garnero E (1998) A correlation between ultra-low basal velocities in the mantle and hot spots. *Science* 281: 546–549.

Wilson JT (1963) A possible origin of the Hawaiian Islands. *Canadian Journal of Physics* 41: 863–870.

Wilson JT (1973) Mantle plumes and plate motions. *Tectonophysics* 19: 149–164.

Wolfe C, Bjarnason IT, VanDecar JC, and Solomon SC (1997) Seismic structure of the Iceland mantle plume. *Nature* 385: 245–247.

Wolfe CJ, Solomon SC, Laske G, et al. (2009) Mantle shear-wave velocity structure beneath the hawaiian hot spot. *Science* 326: 1388–1390.

Wolfe CJ, Solomon SC, Laske G, et al. (2011) Mantle P-wave velocity structure beneath the Hawaiian hotspot. *Earth and Planetary Science Letters* 303: 267–280.

Wolfe CJ, Solomon SC, Silver PG, VanDecar JC, and Russo RM (2002) Inversion of body- wave delay times for mantle structure beneath the Hawaiian Islands: Results from the PELENET experiment. *Earth and Planetary Science Letters* 198: 129–145.

Wolff JA, Ramos FC, Hart GL, Patterson JD, and Brandon AD (2008) Columbia River flood basalts from a centralized crustal magmatic system. *Nature Geoscience* 1: 177–180.

Xu J-F, Suzuki K, Xu Y-G, Mei H-J, and Li J (2007) Os, Pb, and Nd isotope geochemistry of the Permian Emeishan continental flood basalts: Insights into the source of a large igneous province. *Geochimica et Cosmochimica Acta* 71: 2104–2119.

Xu W, Lithgow-Bertelloni C, Stixrude L, and Ritsema J (2008) The effect of bulk composition and temperature on mantle seismic structure. *Earth and Planetary Science Letters* 275: 70–79.

Xu Y, Chung S-L, Jahn B-M, and Wu G (2001) Petrologic and geochemical constraints on the petrogenesis of Permian-Triassic Emeishan flood basalts in southwestern China. *Lithos* 58: 145–168.

Xu Y, He B, Chung S-L, Menzies MA, and Frey FA (2004) Geologic, geochemical, and geophysical consequences of plume involvement in the Emeishan flood-basalt province. *Geology* 32: 917–920.

Xue M and Allen RM (2005) Asthenospheric channeling of the Icelandic upwelling: Evidence from seismic anisotropy. *Earth and Planetary Science Letters* 235: 167–182.

Yale MM and Phipps Morgan J (1998) Asthenosphere flow model of hotspot-ridge interactions: A comparison of Iceland and Kerguelen. *Earth and Planetary Science Letters* 161: 45–56.

Yamamoto M, Morgan JP, and Morgan WJ (2007) Global plume-fed asthenosphere flow-II: Application to the geochemical segmentation of mid-ocean ridges. In: Foulger GR and Jurdy DM (eds.) *Plates, Plumes, and Planetary Processes, GSA Special Paper 430:* Geological Society of America.

Yang T and Shen Y (2005) P-wave velocity structure of the crust and uppermost mantle beneath Iceland from local earthquake tomography. *Earth and Planetary Science Letters* 235: 597–609.

Yang T, Shen Y, van der Lee S, Solomon SC, and Hung S-H (2006) Upper mantle structure beneath the Azores hotspot from finite-frequency seismic tomography. *Earth and Planetary Science Letters* 250: 11–26.

Yasuda A and Fujii T (1994) Melting phase relations of an anhydrous mid-ocean ridge basalt from 3 to 20 GPa: Implications for the behavior of subducted oceanic crust in the mantle. *Journal of Geophysical Research* 99: 9401–9414.

Yaxley GM and Green DH (1998) Reactions between eclogite and peridotite: Mantle refertilization by subduction of oceanic crust. *Schweizerische Mineralogische und Petrographische Mitteilungen* 78: 243255.

Yuan HY and Dueker K (2005) Teleseismic P-wave tomogram of the Yellowstone plume. *Geophysical Research Letters* 32: L07304. http://dx.doi.org/10.01029/02004GL022056.

Zhang J and Green HW (2007) Experimental investigation of eclogite rheology and its fabrics at high temperature and pressure. *Journal of Metamorphic Geology* 25: 97–115.

Zhang T, Lin J, and Gao J (2011) Interactions between hotspots and the Southwest Indian Ridge during the last 90 Ma: Implications on the formation of oceanic plateaus and intra-plate seamounts. *Science China-Earth Sciences* 54: 1177–1188.

Zhao D (2007) Seismic images under 60 hotspots: Search for mantle plumes. *Gondwana Research* 12: 335–355.

Zheng L, Yang Z, Tong Y, and Yuan W (2010) Magnetostratigraphic constraints on two-stage eruptions of the Emeishan continental flood basalts. *Geochemistry, Geophysics, Geosystems* 11.

Zhong H, Campbell IH, Zhu W-G, et al. (2011) Timing and source constraints on the relationship between mafic and felsic intrusions in the Emeishan large igneous province. *Geochimica et Cosmochimica Acta* 75: 1374–1395.

Zhong H, Zhu W-G, Chu Z-Y, He D-F, and Song X-Y (2007) Shrimp U-Pb zircon geochronology, geochemistry, and Nd-Sr isotopic study of contrasting granites in the Emeishan large igneous province, SW China. *Chemical Geology* 236: 112–133.

Zhong SJ and Watts AB (2002) Constraints on the dynamics of mantle plumes from uplift of the Hawaiian Islands. *Earth and Planetary Science Letters* 203: 105–116.

Zhou H and Dick HJB (2013) Thin crust as evidence for depleted mantle supporting the Marion Rise. *Nature* 494: 195–200.

Zhou Z, Malpas J, Song XY, et al. (2002) A temporal link between the Emeishan large igneous province (SW China) and the end-Guadalupian mass extinction. *Earth and Planetary Science Letters* 196: 113–122.

Zindler A and Hart S (1986) Chemical geodynamics. *Annual Review of Earth and Planetary Sciences* 14: 493–571.

Zlotnik S, Afonso JC, Diez P, and Fernandez M (2008) Small-scale gravitational instabilities under the oceans: Implications for the evolution of oceanic lithosphere and its expression in geophysical observables. *Philosophical Magazine* 88: 3197–3217.

This page intentionally left blank

7.11 The Core–Mantle Boundary Region

JW Hernlund, Tokyo Institute of Technology, Tokyo, Japan
AK McNamara, Arizona State University, Tempe, AZ, USA

© 2015 Elsevier B.V. All rights reserved.

7.11.1 Introduction 462
7.11.2 Overview of the CMB Region 462
7.11.2.1 *D″* Discontinuity 463
7.11.2.2 Large Low-Shear-Wave-Velocity Provinces 464
7.11.2.3 Ultralow-Velocity Zones 464
7.11.2.4 Outermost Core Stratification 465
7.11.2.5 Plumes, Slab Ponds and Graveyards, and Other Features 465
7.11.3 *D″* Discontinuities 466
7.11.3.1 Seismic Observations and Inferences 466
7.11.3.1.1 Discontinuities near the top the *D″* layer 466
7.11.3.1.2 Velocity decrease discontinuities 467
7.11.3.2 Models and Mechanisms 468
7.11.3.2.1 Geodynamical predictions 468
7.11.3.2.2 Postperovskite phase transition 468
7.11.3.2.3 Postperovskite double-crossing 469
7.11.3.2.4 Complexities in the postperovskite phase change 469
7.11.3.3 *D″* Discontinuity Summary 471
7.11.4 Large Low-Shear-Wave-Velocity Provinces (LLSVPs) 472
7.11.4.1 Seismic Observations 472
7.11.4.1.1 Seismic tomography 472
7.11.4.1.2 Negative correlation between shear-wave and bulk sound velocities 473
7.11.4.1.3 Sharp gradients along LLSVP margins 473
7.11.4.1.4 Summary of seismic observations 474
7.11.4.2 Hypotheses Regarding the Cause of LLSVPs 474
7.11.4.2.1 Introduction 474
7.11.4.2.2 Significance of understanding the cause of LLSVPs 475
7.11.4.2.3 Thermal hypotheses 475
7.11.4.2.4 Thermochemical hypotheses 477
7.11.4.3 Long-Term Stability of LLSVPs 490
7.11.4.4 Summary and Future Work 491
7.11.5 Ultralow-Velocity Zones (ULVZs) 491
7.11.5.1 Seismic Observations 491
7.11.5.2 Basic Dynamical Principles of a ULVZ Layer 492
7.11.5.3 Mechanisms and Hypotheses 492
7.11.5.3.1 Partial melt 493
7.11.5.3.2 Composition anomaly 493
7.11.5.3.3 Partial melting of compositionally distinct material 494
7.11.5.3.4 Basal magma ocean 494
7.11.5.4 Summary 495
7.11.6 Outermost Core Stratification 495
7.11.6.1 Seismological Constraints on the Stratified Layer 495
7.11.6.2 Constraints from Geomagnetism 496
7.11.6.3 Stability of the Outermost Core Layers 497
7.11.6.4 Composition of the Stratified Layer 498
7.11.6.4.1 Alloy incompatibility from inner-core growth 498
7.11.6.4.2 Core–mantle reaction-induced stratification 499
7.11.6.5 Summary and Perspective 499
7.11.7 Core–Mantle Heat Exchange 499
7.11.7.1 Plume Heat Flow 499
7.11.7.2 CMB Heat Flow Estimates from *D″* Conductivity and Thermal Gradients 500
7.11.7.2.1 Estimates of mantle thermal conductivity 501
7.11.7.2.2 Estimates of temperature change and boundary layer thickness 501
7.11.7.2.3 Geothermal gradients and postperovskite 502

7.11.7.2.4 Geothermal gradient constrained by a double-crossing 502
7.11.7.2.5 CMB heat flux from seismic tomography and boundary layer models 503
7.11.7.2.6 Summary of D'' constraints upon CMB heat flow 504
7.11.7.3 Heat Flow Constrained by Core Evolution and Dynamics 504
7.11.7.3.1 Heat carried by core conduction and convection 505
7.11.7.4 High CMB Heat Flow and Thermal Evolution 506
7.11.8 Core–Mantle Mass Exchange 506
7.11.8.1 Core Versus Mantle Affinity 506
7.11.8.2 Entrainment and Other Direct Core–Mantle Exchange Processes 507
7.11.8.3 Core–Mantle Chemical Reactions 507
7.11.8.3.1 Length scales of equilibration 508
7.11.8.3.2 Disequilibrium-driven reactions 509
7.11.8.3.3 Core saturation? 510
7.11.8.4 Summary 510
7.11.9 Concluding Remarks 511
Acknowledgments 511
References 511

7.11.1 Introduction

The Earth was formed by the accretion of both differentiated and undifferentiated parent bodies, and the large density difference between metallic and ionic components gave rise to internal gravity-driven segregation into the primary regions of our planet: a central metallic core and a rocky oxide mantle. At the current pressure (136 GPa) and temperature (≈4000 K) conditions of the Earth's core–mantle boundary (CMB), the material in the core exhibits a density of 9900 kg m^{-3}, while the density of the mantle is around 5500 kg m^{-3} (Dziewonski and Anderson, 1981), yielding a greater density contrast than that of the Earth's surface. Besides density, the outermost core and lowermost mantle also exhibit strong differences in transport properties such as viscosity (the outer core is liquid, while the mantle is mostly solid), thermal and electrical conductivity, and atomic diffusivity, all of which strongly influence the contrasting dynamics of the core and mantle regions. The strong differences between the mantle and core are reflected in the structure and dynamics of the CMB, where the two major regions of the Earth exchange both heat and matter.

Studying the CMB region is key to building a complete Earth model that integrates the dynamics, structure, and evolution of both major regions of the Earth's interior. By understanding how the myriad dynamics in one region are coupled to those in the other, a broader range of constraints can be brought to bear by linking together major hypotheses related to the evolution of Earth's interior through geologic time. Thus, a major discovery in the deep mantle can carry important implications for the core, and vice versa. However, CMB studies are challenging because of the extreme pressure–temperature (P–T) conditions, which strongly affect material properties and have historically been difficult to mimic in laboratory experiments. Direct observations are also hampered by the difficulty of obtaining reliable seismological and electromagnetic observations that are obscured by heterogeneous structure in the shallower mantle. Nevertheless, the adoption of new technologies for obtaining CMB P–T conditions in the laboratory, the improvement in the reliability of ab initio methods for calculating material behavior at extreme conditions, and the proliferation of digital seismograph networks and methods to process large data sets have ushered in a new era in CMB research that is beginning to shed light on this critically important but enigmatic region of the Earth's deep interior. Geodynamics research is also increasingly playing a key role in integrating the available constraints into models of greater scope and predictive capacity, which is essential for the incorporation of constraints from other geologic sciences (e.g., geochemistry).

CMB research is presently in a period of rapid change, and a single paradigmatic view has yet to become firmly established. This chapter presents the evolution in thought regarding the dynamics of the CMB region and the excitement of new discoveries and models currently under way while also surveying some of the current debates and uncertainties. We begin with an overview followed by an exploration of the primary observed features of the CMB region, including the D'' discontinuities, the large low-shear-wave-velocity provinces (LLSVPs) beneath the Pacific and Africa, the relatively thin ultralow-velocity zones (ULVZs) and other thin basal layers at the CMB itself, and the proposition of an outermost core stratified layer just underneath the CMB, which could play an important role in core–mantle chemical interactions. We then examine the exchange of heat and matter across the CMB, a key area of CMB research that aims to connect the core and mantle systems.

7.11.2 Overview of the CMB Region

Before proceeding with our in-depth discussion, we present a summary of CMB region structures and terminology in order to help orient readers. Indeed, there are numerous acronyms, in addition to varying names for features, which are sometimes connected to specific hypotheses. We thus give a brief history for these features and a discussion on the evolution of their descriptive terminology. Finally, we settle upon the accepted modern usage, although we are sure that the nomenclature will continue to evolve in the future. For visual reference in the following discussion, Figure 1 gives an illustration of the basic seismic observations of the primary features of the CMB region, while Figure 2 shows a typical interpretation of these

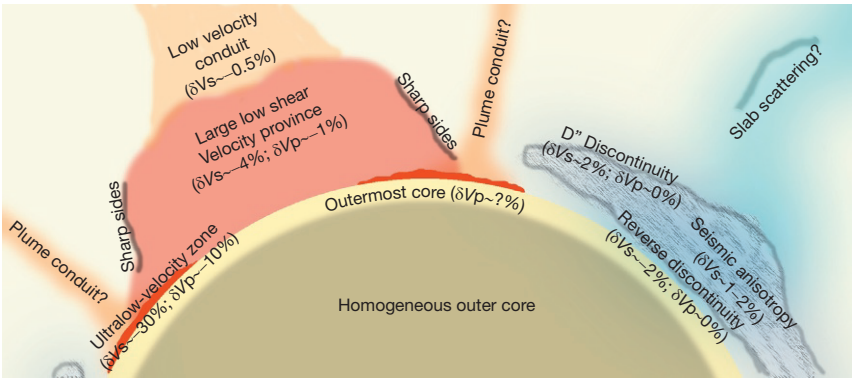


Figure 1 Schematic illustration of some major features of the CMB region proposed by seismological observations. Refer to the text for a more detailed discussion.

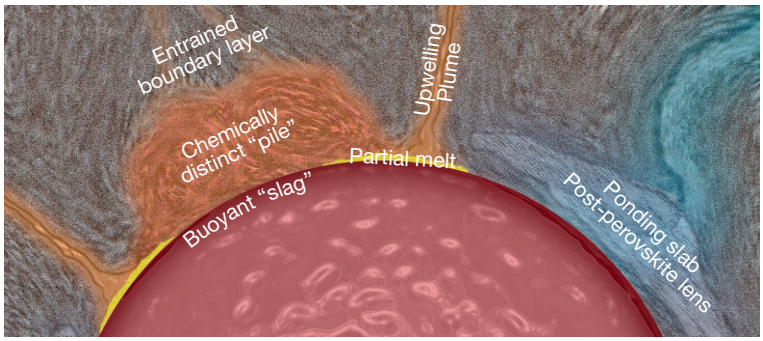


Figure 2 Schematic illustration of a possible interpretation of structures in the CMB region.

features. In each subsection in the succeeding text, the basic observations and the common interpretations listed in Figures 1 and 2 are briefly explained.

7.11.2.1 *D″* Discontinuity

It is important to note from the beginning that the term *D″* has its origin in the alphabetic description of regions in Earth's interior used in the early days of seismology and was meant to describe the anomalous region in the lowermost several hundred kilometers of the mantle (e.g., Bullen, 1949). As such, *D″* is a layer of the Earth's mantle, comprising the lower boundary layer of the mantle side of the CMB transition. The *D″* discontinuity, on the other hand, was originally discovered as a sudden increase in shear-wave velocity occurring several hundred kilometers above the CMB many years later (Lay and Helmberger, 1983). Further work has found that the discontinuity may not be globally ubiquitous or may have significant depth variations that make it difficult to detect in some regions (see reviews by Lay et al. (2004) and Wysession et al. (1998)). Therefore, the *D″* layer should not be thought of as a region that is simply defined by a seismic discontinuity, in the same manner as shallower regions of Earth's mantle are separated by clear global seismic-wave speed jumps. Where it exists, the *D″* discontinuity is revealed by a triplication in the arrival times of S-waves turning in the lowermost several hundred kilometers of the Earth's mantle, clearly producing reflected, refracted, and penetrating branches in the travel-time curve. In some cases, *P*-wave discontinuities have also been reported, although it is not clear if they are correlated with the same *S* discontinuities, and they also exhibit greater variations in both the strength and sign of the velocity jump (Wysession et al., 1998).

The cause of the *D″* discontinuity was a mystery for more than two decades; however, recent mineral physics discoveries appear to have resolved many questions relating to its origin, although numerous details remain to be ironed out. In fact, the present explanation for the *D″* discontinuity as the result of a solid–solid phase change was predicted in advance of its discovery (Sidorin et al., 1998), based on the simple reasoning that the only mechanism that could elevate the discontinuity in seismically fast regions (i.e., beneath cool downwellings) is a strongly exothermic phase transition with a Clapeyron slope of >6 MPa K^{-1} (compositional or thermal frontiers would instead be depressed in such a context). The discovery of a phase change from $MgSiO_3$ perovskite (Pv) to postperovskite (PPv) (Murakami et al., 2004) at the appropriate pressures and with a suitably large transition Clapeyron slope (Oganov and Ono, 2004; Tsuchiya et al., 2004) fits very well with the predictions and remains the most commonly invoked explanation for the *D″* discontinuity at the present time. The PPv phase exhibits a significant degree of elastic anisotropy and should also help to explain observations of seismic anisotropy in these regions (Oganov et al., 2005; Stackhouse et al., 2005). The Pv–PPv phase change may also explain deeper seismic discontinuities in *D″* and suggests that PPv occurs in lens-shape structures rather than a global layer, as a consequence

of the 'double-crossing,' with important implications for characterizing thermal gradients at the base of the mantle (Hernlund et al., 2005).

7.11.2.2 Large Low-Shear-Wave-Velocity Provinces

Since the dawn of seismic tomography, it has been known that the lowermost 1000 km of the mantle contains large-scale 'degree-2' anomalies in seismic-wave speed that are correlated with similar variations in the geoid and that might be simply explained by the presence of abnormally dense matter beneath the Pacific and Africa (Dziewonski et al., 1977). Others have interpreted these features as being large thermal plumes, in which case the terminology is usually 'superplumes' (particularly in a geologic context, e.g., Larson (1991)), but sometimes, the term 'megaplumes' is used in reference to dynamical models (e.g., Matyska et al., 1994; Thompson and Tackley, 1998). It has also been suggested that these regions could be a collection of small plumes, in which case the term used to describe these features is 'plume clusters' (e.g., Schubert et al., 2004). Many lines of evidence suggest that these regions are compositionally distinct from the surrounding mantle, usually based upon evidence such as distinct *S*- and *P*-wave speed variations (e.g., Bolton and Masters, 2001; Hernlund and Houser, 2008; Houser et al., 2008; Lekic et al., 2012; Romanowicz, 2003), sharp gradients and/or reflections of seismic waves from the edges of these features (e.g., Ford et al., 2006b; Luo et al., 2001; Sun et al., 2007a; Wang and Wen, 2004; Wen, 2001), and corresponding variations in both seismic-wave speeds and gravity (e.g., Ishii and Tromp, 1999; Simmons et al., 2006; Trampert et al., 2004). Under the presumption of a compositionally distinct origin, these features have been described by terms such as 'cryptocontinents' (Stacey, 1992), 'domes' (e.g., Davaille, 1999), and 'piles' (e.g., McNamara and Zhong, 2004), some of these terms having further implied meaning with regard to their dynamical manifestation. In reference to the high bulk modulus that can account for their distinct *S*- and *P*-wave speed characteristics, dynamical models attributing chemically distinct dense structures with a corresponding anomalous compressibility also refer to them as 'high-bulk modulus structures' (HBMS) (Tan et al., 2011). In order to establish a purely descriptive term based on seismic observations, while avoiding explicit reference to a particular origin hypothesis, Thorne Lay suggested the terminology LLSVPs (Lay, 2007), which appears to have become more commonly used in recent years than previous terminology. Thus, we adopt the acronym LLSVPs to refer to these large laterally discontinuous regions exhibiting low seismic-wave speeds residing beneath the Pacific and Africa.

Whatever their name, it is clear that LLSVPs are of central importance in deciphering the dynamics of the CMB region. Many active hot spots attributed to deep-seated plumes appear to be located near the edges of LLSVPs (Thorne et al., 2004), and plate motion reconstructions seem to place many of the large igneous provinces (LIPs) in the vicinity of the LLSVPs when they originally erupted in the past 200 My (Torsvik et al., 2006). On the basis of seismic tomography models, it has also been proposed that the LLSVPs channel significant volumes of warm material upward into the shallow mantle in a manner that does not manifest as straightforward hot spots (e.g., Romanowicz and Gung, 2002). Geochemically, the LLSVPs have become a popularly invoked repository for putative 'primitive reservoirs' (e.g., Tolstikhin and Hofmann, 2005) or 'slab graveyards' (e.g., Christensen and Hofmann, 1994), and deciphering their true origin is of utmost importance for models of Earth's chemical and isotopic composition. These features have also been proposed to influence geomagnetic variations and reversals (e.g., Clement, 1991; Costin and Buffett, 2004; Laj et al., 1991) or occasionally produce increased volcanic activity at the surface with significant impacts upon the Earth's paleoclimate (e.g., Caldeira and Rampino, 1991; Condie, 2001), perhaps even playing a major role in mass extinction events (e.g., Courtillot et al., 1996; Keller, 2005; Zhou et al., 2002). Indeed, Courtillot and Olson (2007) suggested a correlation between all of these phenomena, highlighting the central role LLSVPs may play in our planet's evolution, from its surface to the center.

It is also noteworthy that, despite the variety and volume of evidence – both direct and indirect – indicating that LLSVPs are composed of compositionally distinct material with a high bulk modulus, there are still some who argue that these features may be explained solely by temperature variations alone. Usually, these perspectives refer to only a single kind of seismic-wave speed variation in the deep mantle (e.g., Schubert et al., 2004), thus limiting their a priori constraints in a way that would allow for multiple interpretations. On the other hand, recently, the interpretation of distinct *S*- and *P*-wave behaviors inside these regions has itself been challenged as an artifact of certain kinds of travel-time determinations combined with finite frequency propagation effects (Schuberth et al., 2012). This particular interpretation also represents a challenge to the LLSVP nomenclature itself: if the *P*-wave speed anomaly is just as strong as that for *S*-wave speeds, then we could simply have called them low-velocity provinces (LVPs).

7.11.2.3 Ultralow-Velocity Zones

Unlike the LLSVPs described in the preceding text, which are 1000 km scale features of the lowermost mantle, there also exist anomalous patchy 10 km layers at the CMB that exhibit more extreme velocity variations: 10% for *P*-waves and 10–30% for *S*-waves (Garnero and Helmberger, 1996). Similar to the other structures discussed in the preceding text, ULVZs are also not globally ubiquitous but rather seem to be present as distinct patches (Thorne and Garnero, 2004). It is also possible that the usual ULVZs are the much thicker manifestation of a layer that is much thinner ($\sim$1 km or less) in other locations (Ross et al., 2004). In some particular regions, it has been possible to employ the *ScP* seismic phase to obtain constraints on ULVZ density, suggesting an $\approx$10% increase relative to the overlying mantle (Rost et al., 2006). Thus, the variations in velocity and density in ULVZ of order 10% are large, comparable with the variations across the Mohorivičić discontinuity at the base of the Earth's similarly thin and variable crust. However, these anomalies are still relatively small in comparison with the density and velocity changes across the CMB itself.

As the veneer separating the Earth's core and mantle, ULVZs are undoubtedly important for understanding the chemical disequilibrium between the mantle and core and the possibility of mass exchange between Earth's major interior regions.

The sharpness of the upper boundary of ULVZs, geographic correlation with mantle upwellings (Williams et al., 1998), and relatively larger *S*-wave velocity reduction relative to *P* motivate a partial melt hypothesis for their origin (Berryman, 2000; Hier-Majumder, 2008; Williams and Garnero, 1996). However, other mechanisms for producing ULVZ have also been considered, such as subduction and gravitational settling of banded-iron formations (Dobson and Brodholt, 2005), sediments floating to the top of an alloy-saturated core (Buffett et al., 2000; Manga and Jeanloz, 1996), strong iron enrichment in PPv (Mao et al., 2004) or magnesiowüstite (Mw) (Wicks et al., 2010), and upward entrainment of iron by morphological instabilities (Otsuka and Karato, 2012) or even poroelastic mechanisms (Petford et al., 2005). However, given the association of all these mechanisms with an enrichment in iron, which is thought to lower the melting temperature and the inevitability of significantly higher CMB temperatures in the Earth's past, these model spaces probably contain significant overlap and may anyways be partially molten at the present or past conditions of the CMB. Recent models that incorporate energy constraints for the geodynamo and corresponding secular cooling of the CMB suggest that ULVZs are the residue of a much larger magma body in the Earth's past, called the 'basal magma ocean' or BMO, which might also help to explain the relationship between ULVZ and LLSVP (Labrosse et al., 2007; Nomura et al., 2011).

7.11.2.4 Outermost Core Stratification

On the underside of the CMB lies the outermost molten metallic core, a region of dramatically different physical and chemical properties relative to the overlying mantle. Outer-core convection is generally thought to produce the Earth's geomagnetic field via dynamo action in electrically conducting metallic fluid (e.g., Braginsky and Roberts, 1995; see also Volume 8), and indeed, most of the Earth's outer core is well fit by an isentropic compression profile consistent with vigorous convection (e.g., PREM; Dziewonski and Anderson, 1981). The presence of convective motions of order 10^{-4} m s^{-1} atop the core can be inferred from geomagnetic secular variations, providing important constraints on the dynamical regime of the core (e.g., Bloxham and Jackson, 1991). However, it is not clear that coherent radial convective motions penetrate all the way up from the bulk of the outer core to the CMB, and it has long been suspected that the outermost ~100 km of the core may be gravitationally stratified owing to excess accumulation of light alloying elements (e.g., Fearn and Loper, 1981) and/or a subadiabatic thermal gradient (e.g., Labrosse et al., 1997). The enrichment of incompatible light elements at depth owing to inner-core growth either could contribute to formation of such a stratified layer (Fearn and Loper, 1981) or could work to disrupt its accumulation (e.g., Loper, 1978). On the basis of improved models for core–mantle chemical equilibria (e.g., Asahara et al., 2007; Frost et al., 2010; Ozawa et al., 2008), it has also been proposed that such a layer could form as a consequence of core–mantle reactions that cause excessive accumulation of oxygen atop the core (Buffett and Seagle, 2010). Given the low viscosity of the outer core (e.g., de Wijs et al., 1998) and inefficiency of viscous entrainment, the only processes capable of mixing such a stratified layer downward involve turbulent entrainment and/or thermal diffusion. Departures in density sufficiently large to be seismically detectable are more than sufficient to stabilize such a layer over geologic timescales (Buffett and Seagle, 2010).

While some features of the geomagnetic secular evolution may be consistent with a stratified layer atop the core (Whaler, 1980), so long as it is not much thicker than ~100 km (Gubbins, 2007), these interpretations are inherently nonunique, and therefore, we must also rely upon seismological verification of the stratification hypothesis. Unfortunately, the only seismic phases capable of probing the outermost core for such a stratified layer (i.e., Sm*KS*) are very sensitive to the strong heterogeneity at the base of the mantle (Garnero et al., 1993a). Many seismological models have proposed a velocity decrease in the uppermost core relative to PREM (Eaton and Kendall, 2006; Helffrich and Kaneshima, 2010; Tanaka, 2004, 2007); however, the models exhibit significant differences, and some authors have since argued for a lack of evidence for stratification (Alexandrakis and Eaton, 2010). Nevertheless, it is clear that the presence of any stratified layer has important connections to dynamics in the CMB region, owing to the sensitivity of such structures to heat flow and potential chemical reactions. Furthermore, the presence of a stratified layer may render some core–mantle chemical exchange mechanisms implausible, such as the upward transport of fractionated osmium isotopes (Brandon and Walker, 2005) or material exhibiting excess Fe/Mn (Humayun et al., 2004), since transport across such a layer would be limited to diffusion alone.

7.11.2.5 Plumes, Slab Ponds and Graveyards, and Other Features

There is presently little doubt that at least some subducted lithosphere sinks to the very bottom of the mantle, consistent with basic observations such as the geoid above subduction regions (Hager, 1984) and seismological imaging that reveals high-seismic-wave speed features extending to the bottom of the mantle in portions of the circum-Pacific region (e.g., Grand, 1994, 2002). There is also a good correlation between the expected locations of subducted slabs and regions of high seismic-wave speed in the D'' layer (Lithgow-Bertelloni and Richards, 1998; Ricard et al., 1993). The presence of slabs may also be essential for stabilizing postperovskite in the deep mantle consistent with seismological observations (e.g., Hernlund and Labrosse, 2007). For these reasons, among others, the focus of dynamical studies has turned to questions regarding what happens to subducted lithosphere when it reaches the D'' layer, its relation to some of the key features of the CMB region discussed in the preceding text, and the possibility of 'slab graveyards' where ancient crust is sequestered for over relatively long residence times.

There also exist a variety of other features in the D'' region that are key to dynamical interpretations. Morgan (1971) proposed that LIPs and long-lived hot spots at the Earth's surface are caused by plumes rising from the deep mantle and assume the form of large plume heads followed by trailing conduits of hot buoyant rock. Additionally, ocean island basalts (OIBs) produced by hot spots carry unique isotopic and compositional signatures that are distinct from the mid-ocean ridge basalts (MORBs) that sample large swaths of the upper mantle,

suggesting a deeper origin for OIBs (Hofmann, 1997). Some seismological analyses claim to have sufficient resolving power to image these plumes, in support of this hypothesis (e.g., Montelli et al., 2004). Hotspots exhibit topographic 'swells' that can be related to the flux of buoyant material from below (Davies, 1988; Sleep, 1990), and when assigned a temperature contrast consistent with petrologic inferences (e.g., Takahashi et al., 1993), these account for a minimum heat flow of 2–4 TW, slightly less than 10% of the total surface heat flow. Note that this number is sometimes mistakenly adopted as the CMB heat flow, but many studies indicate that the CMB heat flow should be much larger than plume heat flow measured in the upper mantle, although the precise reasons are still debated (Labrosse, 2002; Mittelstaedt and Tackley, 2005; Zhong, 2006). When corrected for these effects, in addition to adiabatic heating/cooling upon compression/decompression, dynamical models suggest that the implied heat flow carried by plumes from the deepest mantle consistent with hot spot swell topography exceeds 10 TW, consistent with more recent inferences of CMB heat flow that will be discussed in detail in later sections.

7.11.3 *D″* Discontinuities

One of the most prominent features of the D'' layer is discontinuous variations in elastic-wave speeds, which can be detected by the analysis of variations in seismic waves that pass through the deep Earth. The most well-known and characterized discontinuity is an increase in the propagation speed of shear (S) waves near the top of the D'' layer, 200–400 km above the CMB (see Wysession et al. (1998) for a review). This particular discontinuity has typically been referred to as *the* D'' discontinuity; however, in recent years, other kinds of discontinuities have also been reported, which, although less well characterized, may be very important for revealing further details and dynamics of the mantle side of the CMB region. Additionally, a discontinuity near the top of D'' is not observed everywhere and may be either difficult to detect or genuinely absent in some portions of the deepest mantle. Thus, the D'' layer is not itself defined by the presence of such a discontinuity.

In this section, we will discuss the main types and features of the known D'' discontinuities, recent interpretations of the cause of discontinuities as a consequence of a perovskit to postperovskite phase transition, and the important dynamical implications of these various scenarios. Relative to other main features of the CMB region, many features of the D'' discontinuities are known with a fair degree of accuracy, and a simple interpretation for producing most of these has already become widely accepted. However, it is less clear that all discontinuities can be produced by the same mechanism, thus indicating the need for other mechanisms to explain all of these features. In relation with the D'' discontinuity, we also discuss the strong anisotropy that is often associated with these kinds of discontinuities, which should provide important constraints on the dynamics of the lowermost mantle.

7.11.3.1 Seismic Observations and Inferences

In this section, we summarize some of the primary seismic observations regarding D'' discontinuities. Because much of this subject has already been well reviewed elsewhere (e.g., Lay, 2007; Lay et al., 1998, 2004; Wysession et al., 1998), we give only a basic summary of the observations while emphasizing those that have an important bearing upon the dynamics of the CMB region.

7.11.3.1.1 Discontinuities near the top the D″ layer

It is important to understand why the discontinuity in S-wave speed near the top of D'' is a well-established feature of some regions in the deep mantle. A rapid increase in seismic velocity with depth can be detected unambiguously by the analysis of seismic travel times for waves turning above and below the depth of its occurrence. The basic scenario is described in **Figure 3**, showing ray paths that turn above the discontinuity, waves that turn within the gradient region (diffraction waves), and waves that turn below the discontinuity. Observation of such a triplication is usually regarded as an unambiguous indication of a seismic-wave speed increase with depth. Lay and Helmberger (1983) established the occurrence of such a wave field triplication for horizontally polarized shear waves turning above the CMB, corresponding to a 2–3% increase in velocity about ~280 km above the CMB. This S discontinuity, where it is observed, may have a gradient thickness of up to ~75 km (Wysession et al., 1998).

Many attempts have been made to correlate the S discontinuity with other features of the deep mantle. Observations of P-wave discontinuities have also been reported near the top of D'' (Wright et al., 1985); however, they are not clearly correlated with – and are more variable than – the S-wave discontinuities (Lay et al., 1998; Wysession et al., 1998). Another main feature of D'' seismic structure is the presence of variable anisotropy (Mitchell and Helmberger, 1973). S discontinuities

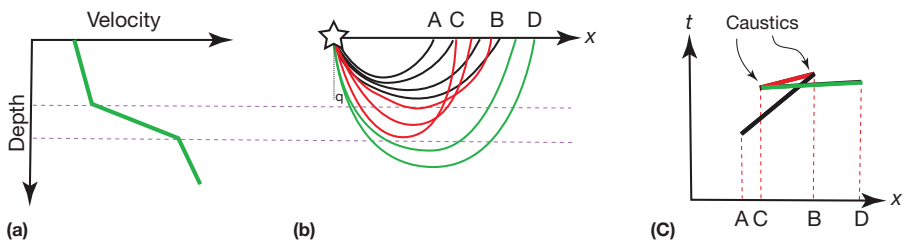


Figure 3 Illustration of how a seismic velocity increase with depth (a) causes ray paths for waves turning above, at, and below the increase (b) to produce a 'triplication' pattern in the travel-time curve. In this figure, x is the distance from the source (star) and t is the travel time of the seismic wave. Base image courtesy of Edward J. Garnero.

near the top of D'' are spatially correlated with the onset of seismic anisotropy in the circum-Pacific (Lay et al., 1998), a region lacking P discontinuities and exhibiting higher-than-average seismic velocities (Figure 4). The anisotropy revealed by techniques such as splitting of S-waves suggests that horizontally polarized S-waves may travel up to several percent faster than vertically polarized S-waves (Lay et al., 1998), a pattern that also appears in more recent global inversions for S anisotropy (Panning and Romanowicz, 2006). Because the anisotropy exhibits similar strength as the S discontinuity atop D'', much or perhaps all of the S-wave speed variations in these regions may be characterized as a discontinuity in horizontally polarized shear waves and only a weak or absent discontinuity for vertically polarized waves. The correlation between strong anisotropy and the D'' discontinuity should provide important clues for assessing the origin of both features.

Note that while D'' discontinuities are typically found in seismically fast regions of the lowermost mantle, there are also some observations of S discontinuities in slow regions, such as the Pacific LLSVP (e.g., Avants et al., 2006; Garnero et al., 1993b). However, in these cases, there is no clear association with the kind of shear-wave anisotropy like that observed in the circum-Pacific region; thus, it has been suggested that slow region D'' discontinuities have an origin distinct from those in fast regions (Lay et al., 1998). Also, the association with P-wave anomalies appears to be different in some regions (Cobden and Thomas, 2013). In addition to these semihorizontal features, nearly vertical discontinuities with gradient thickness less than 100 km have also been detected around the edges of some LLSVPs (Ford et al., 2006b; Luo et al., 2001; Sun et al., 2007a; Wang and Wen, 2004; Wen, 2001), perhaps demarcating the edge of these features, as will be discussed in later sections. It has also been proposed that D'' discontinuities may manifest as a transition from weak to strong scattering atop the D'' layer (Cormier, 2000). The variety of observations and interpretations are therefore not always as simple as a simple S increase discontinuity and weak or absent P discontinuity, suggesting that the features in the deep mantle derive from a variety of origins (Figure 5).

7.11.3.1.2 Velocity decrease discontinuities

In addition to velocity increase discontinuities, there also exists evidence for velocity decrease discontinuities deeper inside the D'' layer. Using techniques similar to those employed in exploration seismology, Thomas and coworkers proposed S velocity decrease discontinuities beneath the more well-established increase discontinuities under Eurasia (Thomas et al., 2004b) and the Caribbean (Thomas et al., 2004a), mostly confined to the lowermost ~100 km of the mantle. However, unlike

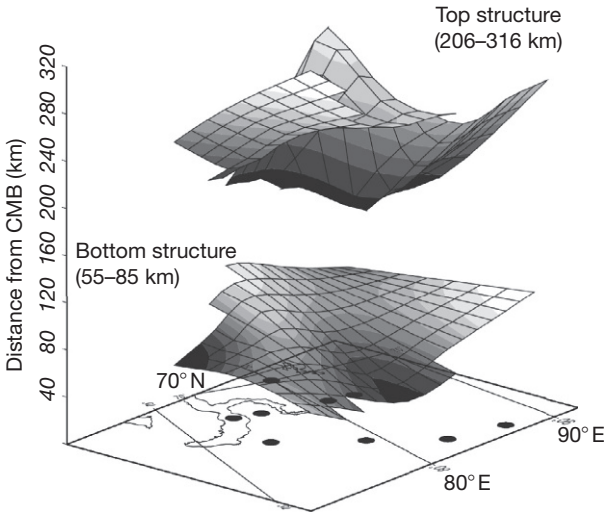


Figure 5 Variation in the depth of an S velocity decrease discontinuity situated beneath the usual S velocity increase discontinuity in the D'' layer under Eurasia. Reproduced from Thomas C, Kendall J, and Lowman J (2004). Lower-mantle seismic discontinuities and the thermal morphology of subducted slabs. *Earth and Planetary Science Letters*, 225: 105–113.

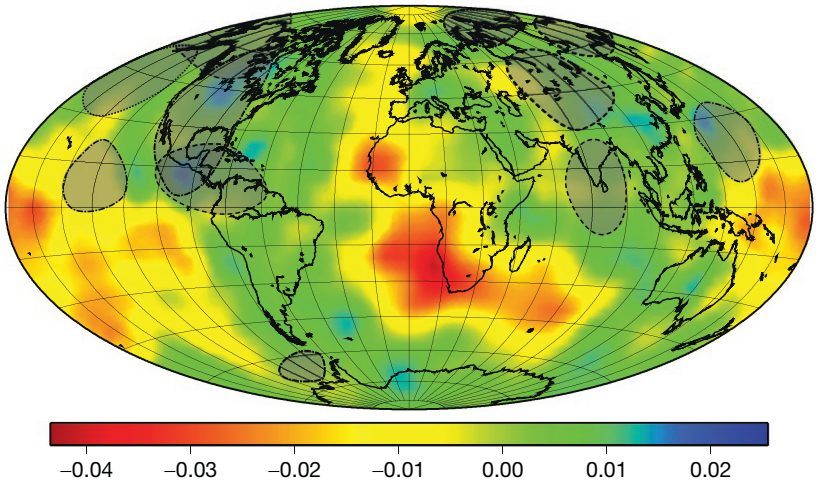


Figure 4 Map showing regions where an S discontinuity near the top of D'' (dot-dashed lines), strong S-wave anisotropy (dotted lines), or both (dashed lines) have been detected (see review by Lay, 2007) along with an addition discontinuity detection under North America (van der Hilst et al., 2007) overlain on an S-wave tomography model at 2800 km depth (Houser et al., 2008). Note that lack of detection does not imply absence, as some regions are not sufficiently covered to provide unambiguous evidence for S discontinuities and/or strong anisotropy.

triplications in the case of velocity increase discontinuities, velocity decrease discontinuities do not yield unambiguous signatures and are more difficult to detect (Flores and Lay, 2005).

The presence of *S* velocity decrease discontinuities beneath *S* velocity increase discontinuities has since been reported using different techniques in a variety of locations, such as waveform inversion of long-period data beneath the Caribbean (Kawai et al., 2007b) and Arctic (Kawai et al., 2007a), waveform modeling beneath the Cocos Plate (Sun et al., 2006), seismic imaging using the 'generalized Radon transform' beneath North America (van der Hilst et al., 2007), and dense array stacking and reflection profiling beneath the central Pacific (Lay et al., 2006) and central America (Hutko et al., 2008).

7.11.3.2 Models and Mechanisms

We now consider proposed mechanisms to explain the presence of discontinuities near the top and within the D'' layer. Prior to 2004, the cause of the *S* discontinuity near the top of D'' was attributed to variations in phase, temperature, and/or composition (Wysession et al., 1998), but the discovery of a perovskite (Pv) to postperovskite (PPv) phase change in $MgSiO_3$ now provides the most plausible explanation (Murakami et al., 2004). The discovery of the Pv–PPv phase change reinvigorated research efforts and may reveal important details about the D'' thermal boundary layer that were previously not well known.

7.11.3.2.1 Geodynamical predictions

Sidorin et al. (1998) predicted a phase change origin for the *S* discontinuity atop D'' 6 years prior to the discovery of a Pv–PPv phase change in $MgSiO_3$, with remarkably similar characteristics. They used thermochemical convection models of deep mantle convection interacting with a phase change to generate synthetic seismograms to test for consistency with observed data. Sidorin et al. first found that thermal gradients were not steep enough to account for the sharpness of observed seismic discontinuities. Compositional frontiers were depressed beneath downwellings and elevated beneath upwellings, which contradicted the observation of discontinuities several hundred kilometers above the CMB in the seismically fast (presumably cool) circum-Pacific region. Only a phase change exhibiting a strongly positive Clapeyron slope was able to elevate the discontinuity in cold regions in a manner consistent with the geography and observed sharpness of D'' discontinuities. They predicted a Clapeyron slope of ~6 MPa K^{-1}, an *S*-wave velocity change of ~1%, and a very weak *P*-wave velocity change. Although no suitable phase change was known at the time, this seminal prediction laid the foundation for the later discovery of the Pv–PPv phase change and facilitated more rapid acceptance of a Pv–PPv origin for the D'' discontinuity.

In the absence of known phase changes, modeling efforts were also undertaken in order to better understand how lower-mantle mineral assemblages could produce the kinds of anisotropy observed in the D'' region, particularly given the association between strong anisotropy and discontinuities in the circum-Pacific. Using a composite rheology, McNamara et al. (2001) used numerical models of mantle convection to investigate the possibility that a transition from diffusion creep to dislocation creep might occur in the lowermost mantle boundary layer and suggested that dislocation creep could become dominant beneath downwelling, while upwellings tend to remain in the diffusion creep regime. If dislocation creep could give rise to lattice-preferred orientation (LPO) beneath downwellings, it may provide an explanation for the occurrence of strong anisotropy in the circum-Pacific (McNamara et al., 2003), a region that is expected to be influenced by subducted oceanic lithosphere, consistent with its higher-than-average seismic velocity and historical locations of subduction zones (Lithgow-Bertelloni and Richards, 1998; Ricard et al., 1993). If the onset of LPO were relatively rapid, then such a mechanism might also produce a D'' discontinuity, in which case it would represent the frontier between diffusion- and dislocation-dominant creeps. However, such mechanisms depend on many complexities involved in rock rheology, which are currently debated. While it is presently thought that deformation and/or transformation textures in PPv dominate lower-mantle anisotropy, it has recently been proposed that PPv may inherit textures from Pv, and vice versa, emphasizing the importance of understanding anisotropy development in both Pv- and PPv-bearing phase assemblages (Dobson et al., 2013). The localization of high stresses at the base of downwelling subducted lithospheric slabs is also relevant for more recent interpretations regarding a PPv-related origin for anisotropy, as will be discussed in the succeeding text.

7.11.3.2.2 Postperovskite phase transition

Given the Fe + Mg/Si ratio of the silicate Earth and the stable crystalline phases over the relevant *P*–*T* range, the phase $(Mg,Fe)SiO_3$ Pv is thought to comprise ~80% of the lower mantle, and therefore, changes from Pv to another solid phase have long been considered as a primary candidate for the D'' discontinuity. Early attention focused on the possibility that Pv could dissociate into component oxides SiO_2 and (Mg,Fe)O (e.g., Meade et al., 1995; Saxena et al., 1996); however, experimental reports of such a phase change were controversial and not faithfully reproduced in later experiments. The discovery of the Pv–PPv transition in $MgSiO_3$ at a pressure and temperature similar to those expected atop the D'' layer, with a Clapeyron slope of 7–9 MPa K^{-1} (Murakami et al., 2004; Oganov and Ono, 2004; Tsuchiya et al., 2004) fits very well with the predictions of Sidorin et al. discussed in the preceding text and was accepted relatively quickly in comparison with the earlier-proposed mechanisms for the D'' discontinuity.

Additionally, the PPv phase is composed of a layered crystal structure with an elastic anisotropy (Oganov et al., 2005; Stackhouse et al., 2005) that may be consistent with correlations between the *S* velocity increase discontinuities atop D'' and the onset of strong anisotropy. Unfortunately, the precise mechanism(s) for generating anisotropy via mantle flow and transformation of Pv to PPv has been controversial (Hernlund, 2013). However, many complexities such as the use of analogue materials and distinction between transformation and deformation fabrics are presently being elucidated (e.g., Miyagi et al., 2011). However, other important behaviors are being discovered, such as inheritance of textures between Pv and PPv during phase changes (Dobson et al., 2013). The analysis of the lowermost mantle anisotropy is still a very active field and

should be important for establishing the connection between deformation and seismic anisotropy in the D'' region.

Recently, much attention has turned to the effects of soluble components in Pv and PPv such as $FeSiO_3$ and Al_2O_3 upon the phase change. Akber-Knutson et al. (2005) predicted that Al_2O_3 would cause a significant broadening of the two-phase coexistence region between Pv and PPv, consistent with later experimental inferences (e.g., Andrault et al., 2010; Catalli et al., 2009; Grocholski et al., 2012). The influence of an expected $FeSiO_3$ component has been somewhat controversial. Early results suggested that the addition of a $FeSiO_3$ component helps to stabilize the PPv structure at lower pressures (e.g., Mao et al., 2004; Ono and Oganov, 2005; Spera et al., 2006; Stackhouse et al., 2006). However, later experiments indicated that $FeSiO_3$ should destabilize PPv and shift the transition to higher pressures (Hirose et al., 2006; Tateno et al., 2007). Also, a high-spin to low-spin (e.g., Badro et al., 2004; Stackhouse et al., 2006) or intermediate-spin (e.g., McCammon et al., 2008) electronic transition in iron hosted in Pv and/or PPv may complicate matters, and its role in the phase relations is not completely understood. Partitioning of FeO between Pv–PPv and ferropericlase will also affect the phase diagram and sharpness of any Pv–PPv discontinuity; however, once again, contradictory results have been reported in the literature (e.g., Auzende et al., 2008; Sinmyo et al., 2008). We will return to a discussion regarding the effects of a broad two-phase region in the Pv–PPv transition for seismic observations and models of the D'' in the following sections.

7.11.3.2.3 Postperovskite double-crossing

At the time of discovery of the Pv–PPv phase change in $MgSiO_3$, the range of estimates for the CMB temperature were roughly 4000 ± 500 K, overlapping with the range of estimates of the temperature for the Pv–PPv phase change at CMB pressure. For a constant Grüneisen parameter of 1.5 along a liquid iron isentrope in the outer core (Vočadlo et al., 2003), a CMB temperature of 4000 K corresponds to an inner-core boundary temperature of ~5200 K. If the CMB temperature is greater than that of the Pv–PPv transition, then Pv would be stable at the very bottom of the mantle and PPv could only occur as a layer above the CMB. The PPv-bearing layer would then be bounded above and below by a 'double-crossing' of the phase boundary by the geotherm. Such a scenario is facilitated by the expected curvature of the geotherm in a boundary layer setting, accounting for the gradual transition from advection- to conduction-dominant radial heat transport with depth as one approaches the CMB. For a relatively cool CMB, on the other hand, PPv would be stable all the way to the CMB and only a single-crossing of the geotherm and the phase boundary would be produced. It is important to recognize that the CMB is itself an isothermal surface, exhibiting lateral temperature fluctuations of order 10^{-4} K or smaller and that larger lateral temperature changes would drive core flows (akin to thermal winds) too strong to be compatible with the observed geomagnetic secular variation (Braginsky and Roberts, 1995; Stevenson, 1987). Thus, in the context of processes in the Earth's mantle, which involve lateral temperature changes of order 10^3 K, the CMB is practically isothermal. The temperature of the Pv–PPv transition at CMB pressure, on the other hand, is sensitive to composition variations; thus, lateral variations in the latter could cause the scenario to shift from single-crossing to double-crossing in different compositional settings (Tackley et al., 2007).

The double-crossing predicts a pair of seismic discontinuities at both depths of intersection between the geotherm and Pv–PPv phase boundary where every PPv becomes stable in the D'' layer, one of which is the classical shallower S-wave speed increase discontinuity, in addition to a deeper reversion to Pv, which should induce seismic velocity variations of opposite polarity. A single-crossing, on the other hand, predicts only a single S-wave speed increase discontinuity. The double-crossing can easily explain the absence of D'' discontinuities in some regions, as a consequence of mantle material that is too warm to cross the Pv–PPv phase boundary (Hernlund et al., 2005), whereas the single-crossing model encounters difficulties in 'hiding' the discontinuity in regions where it is not observed, since it predicts a global layer of PPv. Detections of S velocity decrease discontinuities discussed in previous sections (Hutko et al., 2008; Kawai et al., 2007a; Lay et al., 2006; Sun et al., 2006; Thomas et al., 2004a,b; van der Hilst et al., 2007), underlying the shallower S velocity increase discontinuities, support the double-crossing scenario and motivated the original proposal of the double-crossing hypothesis (Hernlund et al., 2005). Integrated geotherm and mineral physics models also show a good agreement between the kind of structures predicted by a double-crossing and both the S- and P-wave characteristics of regions such as the circum-Pacific (Wookey et al., 2005), including very weak negative P discontinuities associated with larger S discontinuities (Cobden and Thomas, 2013; Hutko et al., 2008).

Some of the particular settings where a double-crossing was originally proposed are subject to different interpretations. For example, the imaging of a double-crossing-like structure in the mid-Pacific LLSVP (Lay et al., 2006) may require heat flux in excess of what would be expected in a chemically distinct 'pile' (e.g., Nakagawa and Tackley, 2008). Further illumination of this region by seismic data from USArray reveals that the reflectivity structure is considerably more complex than it initially appeared (Thorne et al., 2013b) and may be better explained by ingestion of MORB crust into piles (Li et al., 2014). There is also evidence for complexity in some other regions of the Pacific that may not be straightforwardly explained by the double-crossing (e.g., Kawai and Tsuchiya, 2009). Thus, while the double-crossing may be a viable model for some regions, such as the circum-Pacific, it does not serve to explain all of the complex structures observed in other regions of the deep mantle.

7.11.3.2.4 Complexities in the postperovskite phase change

The Pv–PPv mechanism for producing a seismic discontinuity structures has been questioned because the transition may become too thick to account for the <75 km gradient thickness consistent with seismic observations (e.g., Andrault et al., 2010; Catalli et al., 2009; Grocholski et al., 2012). However, the two-phase coexistence region width (i.e., the experimentally determined pressure increment) is only strictly an upper bound on the gradient thickness of a discontinuity produced by a phase change. While some authors draw linear trends in phase abundance through the Pv–PPv coexistence region (e.g., Andrault et al., 2010), the actual phase abundance given by the

'lever rule' law of molar species conservation can be highly nonlinear, with important consequences for interpreting these kinds of results. In particular, it has been demonstrated that the effective gradient thickness compatible with seismic observations may be significantly smaller than the two-phase coexistence pressure increment (e.g., Stixrude, 1997). Numerical calculations of a D'' boundary layer geotherm and a divariant Pv–PPv phase diagram with variable pressure increments found that a broadening of the two-phase region could in fact sharpen the gradients in phase abundance at the top and bottom of a PPv lens, opposite to the intuitive picture obtained by distributing the phase change linearly across the two-phase region (Hernlund, 2010). Thus, a broadening of the Pv–PPv coexistence region in P–T space alone does not necessarily present any problems for explaining the observed discontinuities and surprisingly may even help to enhance the sharpness of the discontinuities. However, the extent of this asymmetrical sharpening inside the two-phase region depends on the details of the phase diagram itself, which are still poorly constrained. Additionally, the seismic velocity jump across the sharp portion of the phase change will be smaller than the total velocity jump, perhaps explaining some of the scatter in observed strength of the D'' discontinuity.

Another issue related to the sharpness of a D'' discontinuity is the bulk composition of the rock in which the transition takes place, with the pressure increment observed in diamond-anvil cell experiments being relatively sharp (~3 GPa) for natural olivine, broadening to ~8 GPa for MORB and increasing to as much as ~14 GPa for a 'pyrolite' composition, owing mostly to increasing concentrations of Al_2O_3 (e.g., Andrault et al., 2010; Grocholski et al., 2012). As will be discussed in the next section on LLSVP, there are a variety of evidences to indicate that large-scale chemical heterogeneity exists in D''; thus, the behavior of the Pv–PPv transition could change substantially from one location to another depending on the prevailing bulk chemistry of the rocks. The circum-Pacific context is perhaps the simplest case, given the association of this region with long-lived subduction zones and high seismic-wave velocities. Thus, in the circum-Pacific, D'' discontinuities and strong anisotropy are proposed to arise from a Pv–PPv phase change in ponded subducted lithosphere originally composed of ~5 km thick MORB crust and ~80 km thick depleted harzburgitic mantle lithosphere. Owing to the small (~5%) volume fraction of basalt veneers in subducted slabs and the olivine-dominant lithology of harzburgite, sharp phase transitions in this context are therefore compatible with experimental results. The lack of discontinuities in other settings could be due to a broadening of the discontinuity in more Al_2O_3-rich compositions or due to a thinning of the PPv lens in warmer mantle or perhaps a combination of both mechanisms.

Some studies have argued that the Pv–PPv phase change occurs at pressure ranges that traverse the CMB, at least for iron- and/or aluminum-rich compositions (Andrault et al., 2010; Grocholski et al., 2012). However, these different diamond-anvil cell (DAC) studies also yield different P–T conditions and Clapeyron slopes for the Pv–PPv phase transition itself, with reported Clapeyron slopes varying by more than a factor of 2. This can be partly attributed to the choice of different pressure standards and scales, as opposed to experimental error. In particular, some studies (e.g., Catalli et al., 2009; Grocholski et al., 2012; Mao et al., 2006) use the gold pressure scale of Tsuchiya (2003) that yields relatively small Clapeyron slopes (~5 MPa K^{-1}). Other studies use no high-P–T standard but instead rely upon extrapolations of thermal pressure from proposed equations of state for the sample itself (Andrault et al., 2010). Experiments performed simultaneously with both gold and MgO using the pressure scale of Speziale et al. (2001) imply a significantly larger Clapeyron slope (~12 MPa K^{-1}). Aside from the important differences in the Clapeyron slope, absolute pressure differences between these different standards and approaches exceed 10%, yet each of these studies claims much smaller errors. While the MgO pressure scale is sometimes called 'geophysically consistent' because it yields postspinel transition pressures that better agree with the depth of the 670 km discontinuity, MgO reacts with samples having geologic relevant compositions, and gold is therefore preferred in these kinds of experiments. Nevertheless, P–T inferences obtained using the gold scale can be transformed to the MgO scale, after which the Clapeyron slope inferences from a large variety of studies are in better agreement (Hirose et al., 2006). However, the correct choice of pressure scales at CMB conditions is still not completely resolved, and unfortunately, this uncertainty trades off with models of the CMB region and gives rise to large uncertainties in P–T determinations. For example, the PPv double-crossing fits nicely with the Clapeyron slope given by the Speziale MgO scale but is implausible if the Tsuchiya Au scale is more representative of the deep Earth (Hernlund and Labrosse, 2007).

A mechanism involving PPv plasticity and development of anisotropic fabric may also help to produce and/or sharpen the seismic velocity discontinuities associated with a Pv–PPv phase change. Additionally, the change in wave speeds across a Pv–PPv transition may be too small to account for seismic observations if the mixture is isotropic before and after the transition; thus, the development of anisotropic fabric may be required to explain the basic amplitude of the D'' discontinuity (Murakami et al., 2007). It has been suggested that, once the volume fraction of PPv becomes sufficiently large in the two-phase region, deformation of the host matrix will become controlled by PPv rheology. Experiments on analogue materials suggest that PPv is significantly weaker than Pv (perhaps by as much as an order of magnitude) and therefore may undergo rapid deformation at a critical fraction of PPv, causing a rapid onset of deformation and production of anisotropic fabric (e.g., Hunt et al., 2009; Thomas et al., 2011). This kind of mechanism is also consistent with the large stresses expected at the core of cool downwellings (McNamara et al., 2001) where PPv is expected to form in the circum-Pacific context, and if the LPO-inducing deformation mechanism is appropriate, there will be a corresponding sharp increase in the wave speed for horizontally polarized S-waves (V_{SH}) on top of a weaker and broader increase in isotropic S-wave speed (v_{SH}).

The aforementioned combined mechanisms and effects are illustrated in **Figure 6**, which shows a double-crossing (Hernlund et al., 2005), the effect of asymmetrical phase abundance in a 5 GPa thickness two-phase region calculated self-consistently by the lever rule in the model of Hernlund (2010), the associated isotropic S-wave speed structure it generates (Murakami et al., 2007), and the possible enhancement in V_{SH} at the onset of a rheologically critical fraction of PPv

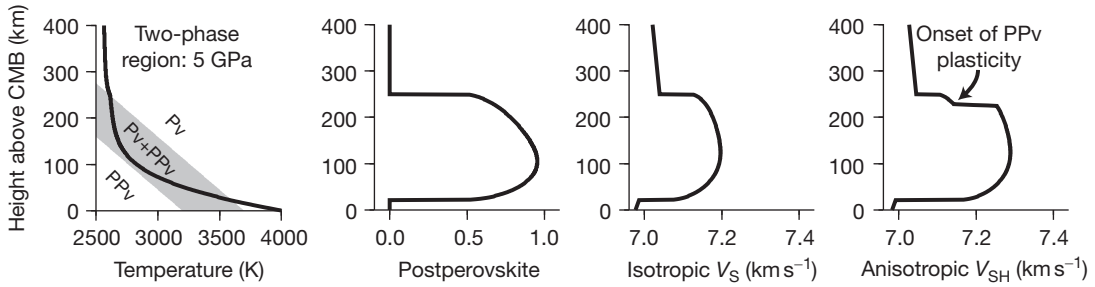


Figure 6 Deep mantle profiles illustrating a potential synthesis of current ideas regarding the occurrence of a Pv–PPv transition and double-crossing inside D'' in a way that could account for broadening of the two-phase region, the sharpness of the discontinuity, a weak change in isotropic S-wave speed (V_S), and the onset of strong seismic anisotropy. The left frame shows the geotherm calculated for interaction with a broad (5 GPa) mixed phase region (gray area), and the second frame from the left shows the resulting postperovskite fraction variation with depth. The profiles are calculated self-consistently using the method of Hernlund (2010). The corresponding isotropic and anisotropic S velocity profiles are shown on the right, in which a sudden onset of PPv plasticity and development of LPO fabric is invoked to explain a sharpening of the upper discontinuity as experienced by SH waves. Note that the transformation from Pv to PPv is never complete in this scenario; there is always a nonzero fraction of Pv phase present throughout the entire depth range. However, structures that are compatible with the PPv double-crossing are nevertheless produced by nonlinear variations in PPv abundance inside the two-phase region.

owing to its weak strength and strong anisotropy (Hunt et al., 2009). The details of this picture may change significantly in the future as the mechanisms are further constrained and/or new mechanisms are considered, and further illumination of D'' by seismic station proliferation will open new windows and perspectives into the complex structure of the CMB region.

7.11.3.3 D'' Discontinuity Summary

Discovery of a Pv–PPv phase transition in $MgSiO_3$ at conditions of the D'' layer revolutionized deep-Earth geophysics and heralded the onset of a new era in which experiments could be routinely performed at the relevant P–T conditions. A Pv–PPv phase change offers the best explanation for the basic first-order features of D'' discontinuities, in addition to their correlation with other features such as strong seismic anisotropy. However, different pressure scales are still applied to DAC experiments at D'' conditions, yielding significant differences in the inferred P–T conditions. And while it is probable that the transition in some bulk compositions occurs over a broad P range, it is also clear that such a measure represents only an upper bound on the gradients in seismic velocity, the latter of which could be very sharp even for very broad transitions.

The relevance of the Pv–PPv phase change for explaining deep mantle structures is probably limited to certain regional and dynamical settings in the lowermost mantle. A Pv–PPv origin for D'' discontinuities is most consistent with the circum-Pacific setting, in which cold ponding slabs are expected to elevate the height of the discontinuity and induce stresses large enough to give rise to anisotropic fabric; positive S discontinuities are associated with weak negative P anomalies; the harzburgite-dominant composition expected for subducted slabs should yield relatively sharp phase transitions; and double-crossing-like structures have been reported by a handful of seismology research groups using a variety of techniques. The mid-Pacific or sub-African lowermost mantle, on the other hand, may be too hot or exhibit thermal gradients too small to be compatible with a Pv–PPv phase change. These LLSVP regions exhibit many features that are probably better explained by variations in bulk composition, and after further examination, structures that once appeared compatible with Pv–PPv origin in the mid-Pacific seem to be more complex and random. However, the phase diagram for Pv–PPv is still constrained poorly enough to permit a variety of exceptional scenarios, and further work should help to limit the number of available options or expand upon the possible range of behavior. Some other regions, such as Eurasia, appear to exhibit a variety of complex behavior, and new kinds of anomalies are still being discovered, thus illustrating that some regions have no straightforward explanation for structures and motivating further work to better characterize them and develop mechanisms compatible with these observations. Increased seismic coverage, particularly in the southern hemisphere, would also help to test hypotheses in a broader array of settings and diversify the underlying observations that motivate research in D''.

One area of research that is very important for interpreting D'' discontinuities, which may yield surprises in the near future, is the development and transformation of anisotropic fabric in the deep mantle. In fact, it may be possible to explain some of the major observations of the D'' layer with fabric changes alone, although most of these must work in concert with a Pv–PPv phase change; thus, the contributions of a variety of factors may yield a variety of possible mechanisms and behaviors to provide a larger tool box for explaining D'' structure. Sorting out the appropriate slip systems for dislocation creep in PPv has proved very difficult, with many groups reporting significantly different results and debates regarding whether some experiments are observing textures due to transformation or deformation (Miyagi et al., 2011). It will ultimately be necessary to integrate these models with dynamical flow models in the D'' region (Nowacki et al., 2013; Walker et al., 2011); however, a consensus view has yet to emerge from all models.

Before concluding the discussion on D'' discontinuities, it is important to note that the Pv–PPv phase change is highly sensitive to variations in temperature, owing to the large Clapeyron slope. Pv–PPv-induced discontinuity depth variations of up to 200 km would correspond to pressure variations of 11 GPa or 850–1400 K lateral temperature variations for a Clapeyron slope of 8–13 MPa K^{-1}. It is difficult to explain such

large lateral temperature anomalies without considering the presence of subducted lithosphere in D'', and therefore, the interpretations of the D'' discontinuity in terms of Pv–PPv are consistent with, and also depend upon, the occurrence of whole mantle convection.

7.11.4 Large Low-Shear-Wave-Velocity Provinces (LLSVPs)

7.11.4.1 Seismic Observations

7.11.4.1.1 Seismic tomography

Over the past two decades, advances in seismic tomography have greatly improved our view of lower-mantle structure (e.g., Antolik et al., 2003; Dziewonski et al., 1977, 2010; Engdahl et al., 1995; Fukao et al., 1992; Grand, 2002; Grand et al., 1997; Gu et al., 2001; Ishii and Tromp, 1999, 2004a; Kuo et al., 2000; Lekic et al., 2012; Li and Romanowicz, 1996; Li et al., 2008; Masters et al., 1996; Megnin and Romanowicz, 2000; Panning and Romanowicz, 2006; Ritsema and van Heijst, 2002; Ritsema et al., 1999, 2004, 2011; Romanowicz and Gung, 2002; Su and Dziewonski, 1997; Takeuchi, 2007; van der Hilst et al., 1997; Zhao, 2004).

Figure 7 displays the seismic tomography model S20RTS (Ritsema and van Heijst, 2002; Ritsema et al., 2004). Higher-than-average seismic velocities are observed beneath the circum-Pacific and Western Asia. Because these faster regions are located beneath zones of ancient subduction, they are generally considered to be caused by thermal anomalies resulting from remnants of subducted lithosphere (e.g., Grand et al., 1997; van der Hilst et al., 1997). This idea is supported by geodynamical models that include the Earth's plate history over the past 120 million years as surface boundary conditions. To the first order, these geodynamical models predict the existence of relatively cool, previously subducted lithosphere descending in the mantle in the same regions that higher-than-average seismic-wave speeds are observed in seismic tomography (Bull et al., 2009, 2010; Bunge et al., 1998; Lithgow-Bertelloni and Richards, 1998; McNamara and Zhong, 2005; Richards and Engebretson, 1992; Ritsema et al., 2007; Schuberth et al., 2009). In most regions, particularly along the Pacific margin of the Americas, seismic tomography reveals subducted lithosphere extending into the

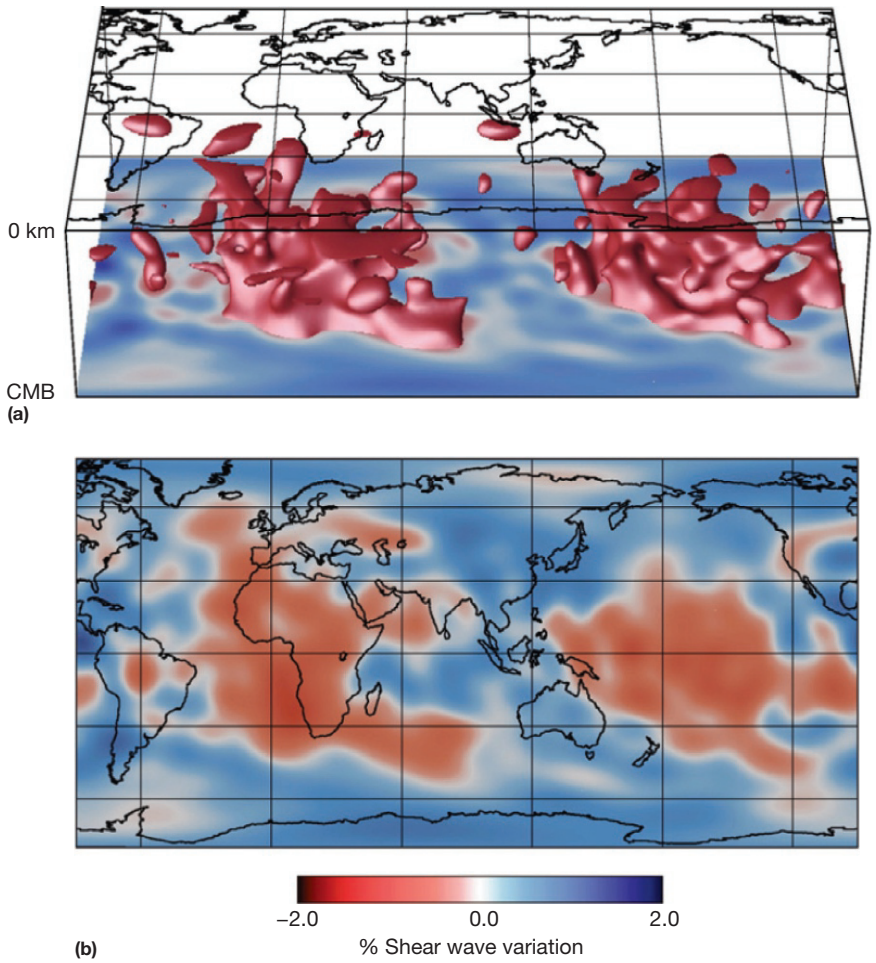


Figure 7 Seismic tomography model S20RTS (Ritsema et al., 1999, 2004). (a) Isosurface representing −0.6% shear-wave velocity anomaly that highlights the presence of large low-shear-wave-velocity provinces (LLSVPs) beneath Africa and the Pacific. (b) Map view of shear-wave velocity variation, slightly above the core–mantle boundary (at 2750 km depth). Reproduced from Bull AL, McNamara AK, and Ritsema J (2009) Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters* 278(3–4): 152–162.

lower mantle (e.g., Grand et al., 1997; Li et al., 2008; van der Hilst et al., 1997), providing strong evidence against a layered mantle, as hypothesized earlier from geochemical observations (e.g., Carlson, 1994; Hofmann, 1997).

The lowermost mantle beneath Africa and the Pacific is characterized by lower-than-average seismic-wave velocities (**Figure 7**). Classically, these slow regions have been called 'superplumes' (e.g., Ni et al., 2002; Romanowicz and Gung, 2002), which has sometimes caused confusion because the name connotes a dynamical dimension to the seismic observation. In recent years, the community has moved toward naming these regions the LLSVPs in order to more carefully separate the observation from the different hypotheses related to their origin. While the name LLSVP refers to shear-wave tomography, these features are also observed in *P*-wave tomography (e.g., Li et al., 2008; Ritsema and van Heijst, 2002; Zhao, 2004). *P*-wave speed anomalies in these features appear to be weaker than for *S* (e.g., see *P* velocity map in **Figure 21** versus the corresponding *S* velocity map in **Figure 4**); however, most *S* models have significantly better resolution than *P* models. In map view of the lowermost mantle, the African LLSVP is a relatively linear feature striking northwest–southeast, and it extends from the southern Indian Ocean, across Africa and the eastern Atlantic, to northwestern Europe. Lekic et al. (2012), using cluster analysis of multiple tomographic models, argued for the existence of a miniature LLSVP in the lowermost mantle beneath Eastern Europe, which they name the 'Perm Anomaly.' It is unclear how this anomaly is related to the larger African LLSVP, but they claim that it is a distinct feature because it is bounded by strong seismic velocity gradients. It is possible that it could be a recently detached portion of the African structure, thus not necessarily distinct in origin. In contrast to the African LLSVP, the Pacific LLSVP appears relatively more rounded in map view and is centered beneath the southwestern portion of the Pacific Ocean. As imaged by seismic tomography, lower-than-average seismic velocities from both LLSVPs extend from the base of the mantle and reach to the upper mantle (e.g., Romanowicz and Gung, 2002). The African LLSVP appears vertically tilted toward the northeast (e.g., Ritsema et al., 1999), whereas the Pacific LLSVP appears to be vertical.

7.11.4.1.2 Negative correlation between shear-wave and bulk sound velocities

Both shear-wave and bulk sound velocities can be computed by joint tomographic inversions for both *P*-wave and *S*-wave data. Bulk sound velocity is a synthetic velocity (computed from both *P*-wave and *S*-wave velocities) that has the practical advantage that it is only sensitive to bulk modulus and density. For a homogenous system of Earth minerals, both shear-wave and bulk sound velocities are expected to systematically vary together as a function of temperature and pressure (e.g., Masters et al., 2000). However, joint tomography studies typically agree that there is a negative correlation between shear-wave and bulk sound velocities in the lowermost mantle, particularly in the lowermost 500 km (e.g., Antolik et al., 2003; Masters et al., 2000; Ritsema and van Heijst, 2002; Su and Dziewonski, 1997; van der Hilst and Karason, 1999). The negative correlation observed in the lowermost mantle is an indication of compositional heterogeneity. Tomography using normal modes (i.e., free oscillations) provides a means to infer density structure. In addition to finding a negative correlation between shear-wave and bulk sound velocities in the lowermost mantle, normal-mode tomography determines that the lowermost mantle regions beneath Africa and the Pacific are characterized by increased relative density (e.g., Ishii and Tromp, 1999, 2001, 2004b). However, the relative density component of normal-mode studies is likely the least constrained (e.g., Lay and Garnero, 2011; Resovsky and Ritzwoller, 1999).

Hernlund and Houser (2008) examined the statistical distribution of areal distribution of *P*- and *S*-wave anomalies as a function of depth in various seismic tomography models. While both *P*-wave and *S*-wave distributions appear Gaussian-like at each depth range, the *S*-wave distributions form a thickened tail on the lower-than-average side of the distribution in the lowermost 700 km of the mantle. They argue that this provides evidence for compositional heterogeneity and/or phase change heterogeneity in the lowermost mantle.

By relating *P*- and *S*-wave velocity and density to changes in temperature and composition (e.g., perovskite and magnesiowüstite content), several probabilistic tomography studies have inferred that tomographic data are better explained with a heterogeneous compositional component in the lowermost mantle (e.g., Deschamps and Trampert, 2003; Trampert et al., 2001, 2004). Deschamps et al. (2012) found that the LLSVPs are better explained by iron and perovskite enrichment than the accumulation of oceanic crust (MORB).

7.11.4.1.3 Sharp gradients along LLSVP margins

Due to the spatially nonuniform distribution of both earthquake sources and seismograph stations, seismic tomography produces an incomplete and blurred view of the Earth's interior. Therefore, the LLSVPs may not accurately represent the true shape of the temperature and/or compositional anomalies that cause them (e.g., Ritsema et al., 2007). Additional seismic studies that rely upon travel-time analysis and waveform modeling have complemented seismic tomography in order to better elucidate the shape of the African and Pacific structures that cause the LLSVPs (e.g., Breger and Romanowicz, 1998; Ford et al., 2006a; He and Wen, 2009, 2012; He et al., 2006; Ni and Helmberger, 2003a,b,c; Ni et al., 2002, 2005; Ritsema et al., 1997; Sun et al., 2007a,b, 2009; To et al., 2005; Wang and Wen, 2004, 2007; Wen, 2001, 2002; Wen et al., 2001). Such studies have found sharp gradients in seismic-wave velocity near the margins of the LLSVPs. These sharp gradients cause a distortion (usually a broadening) of the waveform that appears on a seismogram. By comparing seismic waveforms along an array of seismic stations that are wide enough to include both seismic waves that pass through an anomaly and those that do not, travel-time analysis and waveform modeling can be used to constrain the position of the edges of the anomalies.

Figure 8 provides a summary of recent work that constrains the shape of the anomalies that cause the LLSVPs. Examining the African structure, it is evident that individual studies do not necessarily agree, likely the result of initial assumptions of the seismic model. One must first pose a trial model for the shape and refine it accordingly to find the best fit for the data. The northern extent of the African structure remains

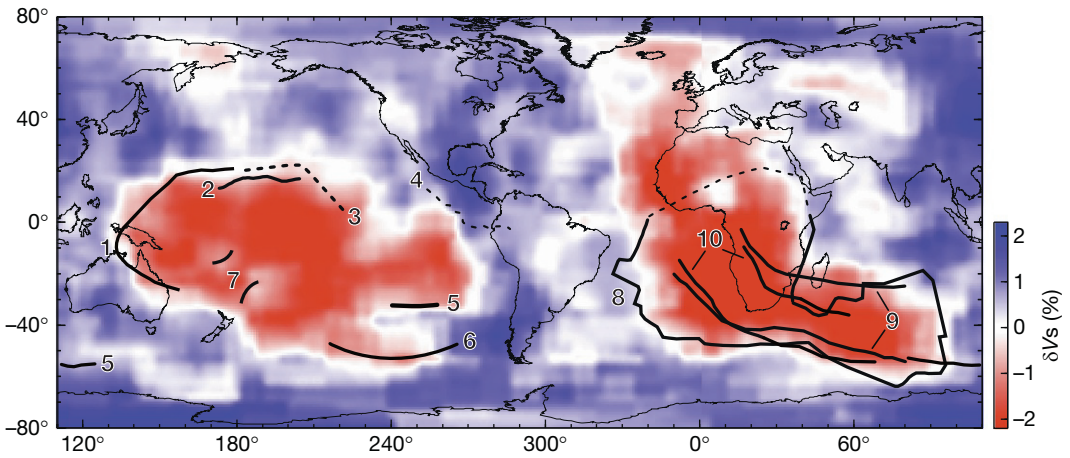


Figure 8 Map showing sharp edges of the large low-shear-wave-velocity provinces as inferred from travel-time and waveform seismic studies. The background map shows shear-wave tomography model TXBW (Grand, 2002). Sharp edges are represented as thick black lines, with dashed lines representing edges inferred from travel times. Numbers indicate the particular seismic studies (1. He et al. (2006); 2. Luo et al. (2001); 3. Breger and Romanowicz (1998); 4. Sun et al. (2007b); 5. To et al. (2005); 6. Ford et al. (2006a); 7. He and Wen (2009); 8. Wang and Wen (2004); 9. Sun et al. (2007b), Sun et al. (2009); and 10. Ni and Helmberger (2003c), Ni and Helmberger (2003a), and Ni and Helmberger (2003b)). This figure was provided by Ed Garnero and Chunpeng Zhao.

unconstrained, and geodynamical models (discussed later) predict that it should extend into Europe and perhaps wrap back down into Eastern Europe (e.g., the 'Perm' anomaly discussed earlier). Much of the Pacific LLSVP margin remains to be mapped. The sharp edges inferred in the middle of the Pacific LLSVP are particularly intriguing and could indicate that multiple anomalies contribute toward creating the LLSVP observed in tomography. He and Wen (2009, 2012) proposed two Pacific anomalies that they name the western Pacific Anomaly and the northern Pacific anomaly. The idea that the Pacific LLSVP is composed of distinct subregions is further supported by the seismic work of Thorne et al. (2013b). As will be discussed later, this is consistent with some geodynamical interpretations.

7.11.4.1.4 Summary of seismic observations

In summary, seismic tomography has revealed a pattern of higher-than-average velocities beneath the circum-Pacific and Asia and large regions of lower-than-average velocity beneath Africa and the Pacific. While the faster regions may be explained as subducted lithosphere extending into the lower mantle, the slower regions beneath Africa and the Pacific are more enigmatic. Joint tomography studies that invert for both *P*-wave and *S*-wave data find that the slower regions are difficult to explain by thermal effects alone and are likely characterized by compositional heterogeneity. Seismic tomography provides an incomplete, blurred view of the interior, so additional seismic methods need to be applied to constrain the shape of the anomalies that cause the LLSVPs. Waveform modeling studies infer sharp velocity gradients near the margins of the LLSVPs, and the anomalies are increasingly being mapped out in better detail. Furthermore, the strong velocity gradients along LLSVP margins are suggestive of sharp thermal and compositional boundaries.

7.11.4.2 Hypotheses Regarding the Cause of LLSVPs

7.11.4.2.1 Introduction

Numerous geodynamical studies have been performed that are geared toward understanding the cause of the LLSVPs and the dynamics associated with them. The various hypotheses that have emerged can be loosely categorized into three groups: thermal, active thermochemical, and passive thermochemical.

The simplest interpretation of the LLSVPs is that they are caused by large thermal anomalies, either as 'megaplumes' (e.g., Tackley, 2000a; Thompson and Tackley, 1998) or as groups of smaller plumes, 'plume clusters' (e.g., Schubert et al., 2004). This interpretation does not necessarily discount the presence of compositional heterogeneity, as inferred from the seismic studies, but it proposes that the dynamics are largely controlled by thermal convection.

Alternatively, other hypotheses propose that lower-mantle dynamics are controlled by a combination of thermal and compositional buoyancy. Mantle convection driven by thermal and compositional buoyancy is typically referred to as 'thermochemical convection'. Thermochemical mantle models include two or more compositional components, each of which is characterized by having a different intrinsic density. The intrinsic density is the density at a particular reference temperature. Thermal variations further modify the density by thermal expansion and contraction. Therefore, the net effective density of mantle material in these models is usually a linear combination of the material's intrinsic density and the density change due to temperature. Thermochemical convection systems that consist of multiple components in which the difference in intrinsic density between components is much greater than the variations in density due to temperature variations tend to develop stratified configurations, which are layered according to intrinsic density. An example of this would be the stratified core–mantle system, in which the

intrinsic density difference between the core and mantle is much greater than density differences due to temperature variation, leading to a layered configuration of the core and mantle. If the Earth's mantle consists of multiple internal compositional components, it is expected that the intrinsic density differences between each component are similar in magnitude to the density differences due to temperature, except for material with very large density anomalies that accumulate at the upper or lower boundaries. Such a system results in a rich variety of possible thermochemical convection behavior (e.g., Tackley, 2007, 2012).

Active thermochemical hypotheses propose that the LLSVPs are caused by compositional anomalies that have an effective density that varies in time such that it fluctuates between being more dense and less dense than the surrounding, background mantle. This type of hypothesis postulates that the LLSVPs are caused by thermochemical structures that cause buoyancy-driven vertical flow. Such structures are typically referred to as 'thermochemical superplumes' (e.g., Davaille, 1999). On the other hand, passive thermochemical hypotheses propose that the LLSVPs are caused by compositional structures that, on average through time, maintain a slightly heavier to near-neutral effective density with respect to the background mantle. Such structures tend to remain at the base of the mantle over geologic time and are passively swept around along the CMB by changing subduction patterns. They do not contribute significant driving force to large-scale mantle convection. Such structures are typically referred to as 'thermochemical piles'.

7.11.4.2.2 Significance of understanding the cause of LLSVPs

Discovering which, if any, of the broadly categorized hypotheses explain the cause of the LLSVPs is a fundamentally important endeavor because each of the conceptual models broadly listed in the preceding text has significantly different consequences toward our understanding of Earth's thermal evolution, temperature structure, chemical evolution, and the driving forces associated with plate tectonics. If the LLSVPs are caused by compositional reservoirs in the lower mantle, these reservoirs will also act as thermal reservoirs, causing the mantle to have a drastically different temperature structure and thermal evolution than that expected for a purely thermally driven mantle (e.g., McNamara and van Keken, 2000; Zhong, 2006). The fundamental ideas associated with the mantle plume hypothesis also hinge upon which of these LLSVP hypotheses may be occurring within the Earth. In particular, geodynamical models consistently reveal that plumes are swept toward upwelling areas governed by the global distribution of subduction (e.g., Zhong et al., 2000). In purely thermal models of mantle convection, plumes are predicted to arise from a thermal boundary layer along the CMB in these regions. In passive thermochemical models however, compositional reservoirs are also driven toward these upwelling regions with mantle plumes forming along their tops (e.g., McNamara and Zhong, 2005). Plumes are predicted to form along the top of upwelling compositional domes in active thermochemical models (e.g., Davaille, 1999). Therefore, our understanding of plume source depends on which hypothesis explains the Earth's mantle. Consequently, using hot spot swell topography to compute plume buoyancy flux to determine core heat loss (e.g., Davies, 1988; Sleep, 1990) is directly affected (e.g., Zhong, 2006).

Discovering the cause of the LLSVPs also directly affects our interpretation of what surface geochemistry tells us about the Earth's interior. The difference in trace element chemistry, particularly incompatible elements and noble gases, between MORBs and OIBs has motivated hypotheses of multiple chemical reservoirs in the Earth's interior including more-primitive mantle material and recycled oceanic crust (e.g., Albarède, 1998; Carlson, 1994; Gonnermann and Mukhopadhyay, 2007, 2009; Hofmann, 1997; Mukhopadhyay, 2012; Tackley, 2000a, 2007; Tolstikhin and Hofmann, 2005; Tolstikhin et al., 2006). Most of these hypotheses rely upon mantle plumes interacting with one or more reservoirs at depth, bringing a small amount of entrained material to surface hot spots as OIBs, whereas divergent boundaries sample depleted background mantle (e.g., Tackley, 2000a). If LLSVPs are indeed caused by compositional reservoirs in the mantle, whether formed by an early differentiation event, the accumulation of oceanic crust, or both, then they are prime candidates to be the geochemical reservoirs hypothesized from geochemical observations.

Finally, compositional reservoirs in the lowermost mantle strongly affects the spatial pattern of heat flux coming out of the core (e.g., Zhang and Zhong, 2011) and could influence the geodynamo (e.g., Driscoll and Olson, 2011; Sreenivasan and Gubbins, 2011).

7.11.4.2.3 Thermal hypotheses

Given that seismic tomography reveals the LLSVPs as having a lower-than-average wave speed, the most obvious hypothesis is that they are caused by hotter-than-average thermal anomalies and are, therefore, megaplumes of the upwelling mantle. However, conventional fluid dynamical calculations at Earth-like convective vigor predict thermal plumes to be much smaller in radius than the LLSVPs (e.g., Olson et al., 1993; Schubert et al., 2004). Thompson and Tackley (1998) found that megaplumes could be formed in axisymmetrical convection models if the temperature dependence of viscosity was sufficiently high (500 kJ mol^{-1} in their case). Their model employed a prescribed initial temperature condition that included a thermal boundary layer at the base of the model. Small-scale instabilities quickly formed within the hot (therefore low-viscosity) lower thermal boundary layer and were swept into a large megaplume along the axis. Given that the experiment was initiated with a prescribed, convection-free (i.e., 1-D) temperature profile, it is unclear whether megaplume formation would regularly repeat over time or if it requires the special case of having an initial boundary layer of uniform thickness. Matyska et al. (1994) suggested that high thermal conductivity in the hottest portions of the D'' layer could also promote the occurrence of large megaplumes, consistent with LLSVP.

Schubert et al. (2004) proposed that the LLSVPs could be caused by poorly imaged clusters of thermal mantle plumes. They provide analytic arguments that demonstrate that megaplumes on the spatial scale of LLSVPs are not dynamically feasible given the rheological and thermal properties of the Earth's mantle. Instead, they hypothesize that numerous smaller plumes form clusters in the large-scale upwelling regions

beneath Africa and the Pacific. This is the result of a highly unstable thermal boundary layer, which nucleates small-scale plume instabilities on timescales short in comparison with the residence time of material at the CMB governed by plate-driven circulation. As viewed by seismic tomography, with its heterogeneous resolution, the clusters of plumes could appear to be grouped together as much larger apparent anomalies.

The plume cluster hypothesis was explored further in McNamara and Zhong (2005), Ritsema et al. (2007), Bull et al. (2009), Schuberth et al. (2009), and Davies et al. (2012). McNamara and Zhong (2005) performed both isochemical and thermochemical convection calculations in 3-D spherical geometry that employed the Earths plate motion history for the past 120 million years as kinematic boundary conditions to guide the historical location of subduction, in a manner similar to Bunge et al. (1998). In the isochemical models, clusters of relatively thin thermal plumes were focused into upwelling regions that formed beneath Africa and the Pacific in response to subduction patterns. In the thermochemical models in which an intrinsically more-dense basal layer was added as an initial condition, subduction history acted to focus the basal layer into piles beneath Africa and the Pacific. Ritsema et al. (2007) converted the temperature fields from these models to variations in shear-wave velocity using a simple linear relationship (e.g., Forte and Mitrovica, 2001). They then applied the seismic resolution operator from tomography model S20RTS (Ritsema et al., 1999, 2004) to the anomalies predicted from the geodynamical models to produce synthetic tomography models. The seismic resolution operator is a measure of how the 'true Earth' is mapped into a seismic tomography model, and when applied to a given distribution of shear-wave velocity anomalies, it shows how the anomalies are distorted by the tomographic inversion. They found that the plume clusters beneath Africa and the Pacific were indeed blurred by the inversion to generate much larger, apparent anomalies. However, these apparent anomalies in the synthetic tomography did not appear as continuous as the LLSVPs observed in actual tomography. Furthermore, the synthetic tomography created from the thermochemical pile models produced a better correlation to actual seismic tomography.

This work was expanded by Bull et al. (2009) who used more thermodynamically consistent conversions from temperature to shear-wave velocity anomaly (e.g., Stixrude and Lithgow-Bertelloni, 2005), explored additional convection parameters, and performed degree-by-degree correlations between synthetic and actual tomography. **Figure 9** displays a representative example of one of the isochemical cases. **Figure 9(a)** displays thermal isosurfaces of the temperature field that highlight the development of plume clusters in the African and Pacific regions. **Figure 9(b)** and **9(c)** shows the temperature field and the

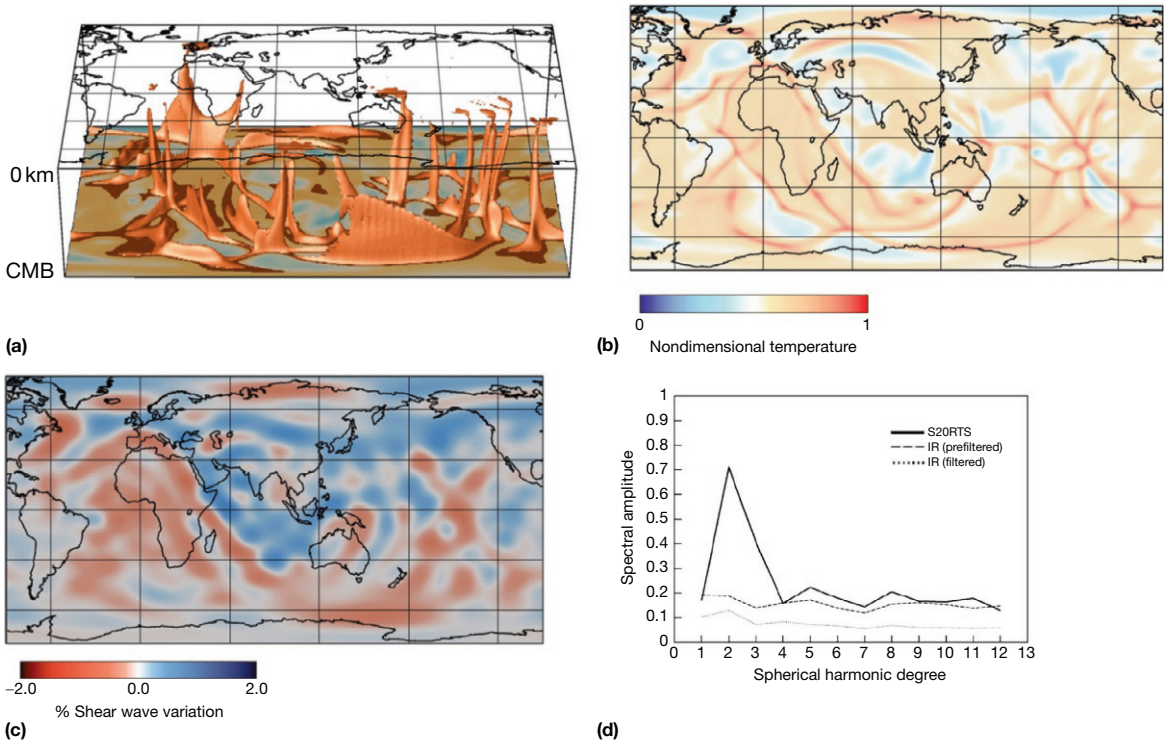


Figure 9 Synthetic shear-wave tomography caused by plume clusters beneath Africa and the Pacific. (a) Plume clusters generated in an isochemical, 3-D spherical convection calculation that utilizes 120 My of plate history as kinematic surface velocity boundary conditions. Results are unwrapped unto a Cartesian box. Isosurfaces of higher-than-average temperature outline the positions of thermal plumes, which form clusters beneath Africa and the Pacific. (b) Map view of the temperature field above the core–mantle boundary (CMB) at 2750 km depth. (c) Map view of synthetic shear-wave tomography generated from the geodynamical model. (d) Amplitude spectra of the spherical harmonic decomposition of the tomographic model S20RTS (Ritsema et al., 1999, 2004) (solid line), temperature field of geodynamical calculation (dark dashed line), and synthetic tomography (light dashed line). Reproduced from Bull AL, McNamara AK, and Ritsema J (2009) Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters* 278(3–4): 152–162, figures 2, 4, and 5.

resultant synthetic tomography in map view at 2750 km depth. From the figures, it is clear that tomography substantially blurs the thermal anomalies, but not enough to generate continuous anomalies as observed in actual tomography (**Figure 7(b)**). **Figure 9(d)** shows the spectral amplitude as a function of harmonic degree for the actual tomography, the prefiltered thermal field, and the synthetic tomography. The synthetic tomography exhibits a relatively flat spectrum that does not exhibit the large degree-2 and degree-3 amplitudes that are characteristic in the actual tomography.

In contrast, Schuberth et al. (2009) performed a similar analysis on isochemical models, but they found a better agreement with seismic shear-wave tomography, showing that thermal models can explain seismic shear-wave tomography. The discrepancy between the studies may be largely attributed to different rheological properties. Bull et al. (2009) employed a temperature-dependent viscosity (10 000× viscosity contrast due to temperature across the model), whereas Schuberth et al. (2009) employed a radial viscosity. Temperature dependence of viscosity may act to produce more numerous, narrower plumes. Davies et al. (2012) is an expansion upon the earlier Schuberth et al. (2009) work, in which they include a mild temperature dependence of viscosity (100× viscosity contrast due to temperature across the model) and compressibility. Their earlier conclusions that isochemical convection can explain the LLSVPs are unchanged, in addition to further predicting strong seismic velocity gradients along the margins.

In summary, if the LLSVPs are caused by thermal-only anomalies, it is likely that they would be clusters of narrow plumes swept into upwelling regions by subduction that resemble larger anomalies when viewed by tomography. Geodynamics-based tomography prediction studies differ on how well plume clusters can resemble LLSVPs and produce sharp seismic gradients along their margins, and these differences are likely due to different model formulations (rheological description and compressibility). Therefore, it is not clear that thermal-only hypotheses for the cause of LLSVPs can be excluded. However, compositional heterogeneity inferred from the negative correlation of shear-wave and bulk sound velocities still needs to be explained. The possibility of plume clusters may also be restricted to a range of material properties that enter into the effective Rayleigh number, such as viscosity, thermal conductivity, and thermal expansivity.

7.11.4.2.4 Thermochemical hypotheses

7.11.4.2.4.1 Early thermochemical studies and ideas

Some of the earliest geodynamical studies on thermochemical convection were motivated by seismic observations that inferred that the D'' region of the lowermost mantle was characterized by strong heterogeneity in both temperature and composition (e.g., Lay, 1987). Davies and Gurnis (1986) and Hansen and Yuen (1988) showed that mantle flow currents could dynamically shape an intrinsically dense layer into a pile at the base of upwelling regions, leading to a thermally and compositionally heterogeneous lowermost mantle. Dense material could be 'leaked' (i.e., entrained) into the upwelling flow depending on the density contrast between the dense and less-dense components (lower density contrast leads to more entrainment).

A possible source for compositional heterogeneity in the mantle is the subduction and eventual segregation (removal from lithosphere) and accumulation of basaltic oceanic crust in the lowermost mantle (e.g., Coltice and Ricard, 1999; Hofmann and White, 1982). The density of basaltic material varies with pressure (e.g., Ringwood and Irifune, 1988), becoming more dense than the surrounding mantle at lower-mantle depths (e.g., Hirose et al., 1999, 2005). Gurnis (1986) was one of the first to explore this hypothesis by performing geodynamical calculations that included a thin layer of tracers at the top of the model to represent basaltic crust. Kinematic velocity boundary conditions were applied to the surface to ensure subduction of this layer into the mantle. Due to computational limitations at the time, these calculations did not have the mesh resolution to resolve the detailed, small-scale dynamics associated with entrainment and the potential segregation of oceanic crust from its lithosphere, so he examined density contrasts representing end-member cases of either full segregation of the crust from the lithosphere (leading to a maximum density contrast) or no separation (leading to a much lower density contrast because crustal density is partially compensated by lower density of subcrustal lithosphere). He found that if the crust does not segregate, it is unlikely to accumulate. If it does segregate, it may accumulate along the bottom of the mantle for fully internally heated cases and beneath upwelling regions for bottom-heated cases. In all cases, the accumulation of the crust is transient and does not fully survive over geologic timescales.

Olson and Kincaid (1991) performed laboratory tank experiments in 3-D Cartesian geometry. One of their experiments consisted of allowing a stratified slab (including dense 'crustal' layer and lighter subcrustal 'lithosphere' layer) to sink within the background mantle. After descending to the bottom of the tank, the slab buckled and formed a pile that underwent segregation by the lighter lithospheric component forming upwelling instabilities, leaving the dense crustal component behind. The dense crustal material spread out along the bottom of the tank in a spoke-like pattern of ridges, presumably outlining the upwelling currents of the larger-scale flow. Afterward, the ridges of dense material became slowly entrained into upwelling thermal plumes.

Whereas the earlier work of Gurnis (1986) employed an isoviscous rheology, Christensen and Hofmann (1994) performed a similar numerical experiment with temperature-dependent rheology and a layered lithosphere (including basaltic crust over a lighter harzburgite layer). They found that about one-sixth of the basalt that subducted could accumulate in piles in the lowermost mantle; however, this varied with density contrast between the basalt and background mantle, temperature dependence of viscosity, and Rayleigh number. The amount of accumulation increased with increased temperature dependence of viscosity and decreased with increasing Rayleigh number (their crustal thickness was tuned to be one-tenth of the upper thermal boundary layer thickness, which decreases with increasing Rayleigh number). Furthermore, they found that the separation of crust from the lithosphere occurs in the lowermost thermal boundary layer (as shown in the Olson and Kincaid (1991) work).

Kellogg and King (1993) examined the formation of a dense mantle layer resulting from percolation of iron and/or

core–mantle reaction products from the core (treated as an effective chemical diffusion). For chemical diffusivities and other material parameters employed in their study, they found that if the dense component is 3–6% more dense than the background mantle, a high-density layer may survive entrainment into the surrounding mantle and persist for long times.

Tackley (1998) is arguably the first comprehensive numerical study investigating thermochemical convection as the cause of the LLSVPs. His model setup involved inserting a more-dense compositional component into an isochemical system that had reached steady-state conditions. He examined Boussinesq and compressible formulations, 2-D and 3-D Cartesian geometries, depth-dependent thermodynamic parameters, depth- and temperature-dependent viscosity, and internal and basal heating. His study focused on using density contrasts that resulted in long-lived, stable thermochemical structures. These structures would later be regarded as thermochemical piles by the community. He found that the more-dense material was swept away from downwellings and accumulated beneath upwelling regions, forming a spoke-like network of ridges. A small amount of entrainment would occur at the base of mantle plumes located at the intersection of these ridges. The piles would always be hotter than the background mantle, even in cases with only internal heating. This is because heat transport from piles to background mantle is limited to conduction. For a range of density contrasts, the more-dense material could form long-lived, isolated structures with a high degree of topography. The range of density contrasts that led to this behavior was dependent upon material parameters and whether compressibility was implemented. Furthermore, the density contrast required for long-term stability decreased with internal heating and temperature-dependent viscosity.

Davaille (1999) performed laboratory tank experiments to explore thermochemical convection, motivated by the observations of elevated surface topography of the African and South Pacific regions (typically referred to as superswells; e.g., McNutt, 1998), the presence of numerous hot spots in these regions (e.g., Duncan and Richards, 1991), and the presence of LLSVPs in the lowermost mantle beneath them. She varied the thickness, density, and viscosity ratios between two compositional materials, initially stratified and isothermal. Results could be categorized into two dynamical regions: doming and stratified, the division of which was primarily caused by density contrast. The doming regime occurred for cases in which the lower layer was less than 1% more dense than the upper layer and consisted of large compositional plumes or domes that would rise and sink in an oscillatory fashion. This oscillatory motion is caused by the alternation of the dominant buoyancy mechanism (i.e., positive thermal buoyancy versus negative compositional buoyancy). The number of oscillations that the domes could experience before becoming completely stirred with the other material depended strongly on the thickness and viscosity ratios. Thermal plumes formed on the top of domes, entraining some of the dome material into them. For experiments in which the lower layer had a density much greater than 1%, the layers remained stratified with a small amount of entrainment occurring between them. These experiments gave rise to the idea that the LLSVPs could be explained by thermochemical domes, which actively rise and sink within the mantle and are a source for thermal plumes that form atop them and slowly entrain them (e.g., Courtillot et al., 2003). The oscillatory domes exhibited by these experiments are commonly referred to as thermochemical superplumes.

Kellogg et al. (1999) performed a 2-D Cartesian numerical thermochemical convection calculation that contained a thick, higher-density basal layer in the lower mantle. The upper layer contained a single convection cell consisting of a thickened downwelling along one boundary. The basal layer was 4% intrinsically more dense than the upper layer; however, because it heated up to high temperature, the effective density increase was less than 1% due to thermal expansion. The layer exhibited a large degree of topography variation, being suppressed beneath the downwelling and elevated beneath the opposing, upwelling region. They demonstrated that increasing the ratio of perovskite to magnesiowüstite (through Si content) could cause the layer to become more transparent to seismic waves. This work demonstrated that an intrinsically dense layer can exist in the mantle and, at the same time, explain tomographic images of some slabs reaching the lowermost mantle.

These pioneering studies on thermochemical convection, combined with increased seismic observations of the lowermost mantle, have collectively motivated our subsequent conceptual models of the mantle. Most current hypotheses suggest that the LLSVPs are caused by the accumulations of compositional heterogeneity, and current questions center upon their dynamical behavior and how they influence large-scale mantle convection, and thermal and chemical transport. Current conceptual models can be very loosely categorized into the sections in the succeeding text.

7.11.4.2.4.2 Passive thermochemical piles of primordial origin

Conceptually, the simplest thermochemical scenario to study involves starting with a stratified system of two layers, an intrinsically more-dense one underlying one that is less dense. The fundamental scientific assumption associated with this type of model is that at some time in the past (presumably early in Earth's history), an initial stratification developed. This could be due to an early differentiation event or process (e.g., Labrosse et al., 2007; Lee et al., 2010; Nomura et al., 2011) or an early accumulation of subducted crust (e.g., Tolstikhin and Hofmann, 2005; Tolstikhin et al., 2006). More generally, multiple processes may have occurred early in the Earth's history to form a 'basal melange' (Tackley, 2012). In this sense, we can consider this type of model as 'primordial,' with the emphasis that a stratified system formed early enough to justify being used as an initial compositional condition in thermochemical modeling. In addition, if the bottom layer material remains effectively more-dense than the other layer over geologic time, it will remain stable (not undergo wholesale overturn) in the lowermost mantle. This condition requires that the negative buoyancy associated with intrinsic density remains greater than the positive buoyancy associated with thermal expansion. Furthermore, if this material remains in contact with the CMB, it will become and remain significantly hotter than the upper layer material, and if rheology is controlled by temperature dependence, it will be weaker than the upper layer material. This weakness can cause the material to be passively shaped and pushed around by larger-scale mantle convection.

This section deals with work that maintains the following assumptions and properties:

1. At some time in early Earth's history, a less-dense layer was underlain by a more-dense layer.
2. Resultant thermochemical structures formed from the originally more-dense layer remain more-dense than the surrounding mantle through geologic time.
3. Rheology is dominated by temperature dependence, such that the hot, more-dense thermochemical structures are weaker than the surrounding mantle and can be passively shaped by larger-scale mantle convection.

Tackley (2002) performed a numerical convection study to investigate three different thermochemical configurations in 3-D Cartesian geometry. One configuration was simply isochemical convection, with no compositional heterogeneity included. The other two involved different volumes of more-dense material, starting in a layered configuration. Cases starting with a more-dense layer equivalent to 30% mantle thickness led to a continuous lower basal layer, similar to that hypothesized by Kellogg et al. (1999), and cases with a more-dense layer equivalent to 10% mantle thickness led to discontinuous thermochemical piles, similar to those of Tackley (1998). For each of these, both constant temperature and zero heat flux bottom thermal boundary conditions were investigated. He compared each thermochemical configuration to the character of seismic heterogeneity as a function of depth. He found that the continuous basal layer configuration led to strong heterogeneity in the mid-lower mantle that is not observed by seismic tomography, whereas the discontinuous thermochemical piles and isochemical cases did a better job, although not conclusively, to describe seismic heterogeneity in the lowermost mantle.

The thermochemical piles in Tackley (2002) developed morphology with characteristics similar to that observed in the thermochemical laboratory experiments of Olson and Kincaid (1991) and would be consistently observed in later laboratory and numerical studies. The more-dense material accumulated into a ubiquitous network of intersecting ridges at the bottom of the mantle, consistent with (and likely governed by) the larger-scale flow occurring in the more-voluminous, less-dense material. The ridges were located at the base of upwelling regions within the less-dense material and were locally elevated where the ridges intersect. Thermal plumes (forming in the less-dense material) were rooted along the ridge tops, particularly at the elevated ridge intersections. More-dense material was entrained into these thermal plumes, providing a mechanism for material exchange across the different materials. Jellinek and Manga (2002, 2004) also found identical morphological features in their Cartesian thermochemical laboratory experiments that involved much thinner dense layers.

An early concern focused on whether the ubiquitous ridge-like network of thermochemical structures observed in the Cartesian thermochemical studies was consistent with the seismic tomography observations of two LLSVPs beneath Africa and the Pacific, each with a somewhat different shape; the African LLSVP is a relatively linear feature, whereas the Pacific one is more rounded. McNamara and Zhong (2004) performed numerical 3-D spherical thermochemical convection calculations to investigate how the morphology of thermochemical structures would be affected by spherical geometry. Each of their cases originated from an intrinsically more-dense layer underlying an intrinsically less-dense one. Various layer thicknesses and density contrasts were investigated for four rheological types: isoviscous, temperature-dependent, temperature- and depth-dependent, and temperature- and composition-dependent. Here, only cases with density contrasts large enough to generate long-lived, stable structures are described. The isoviscous cases resulted in a basal layer with a jagged interface topography that was symmetrical on each side of the interface (i.e., topography looked the same whether viewed from above or below the interface). However, in the temperature-dependent rheology cases, the more-dense material became much weaker than the less-dense material (due to the higher temperature of the more-dense material). It was, therefore, easily swept around and shaped by larger-scale convection currents within the less-dense material. Interestingly, the resultant thermochemical piles developed a morphology effectively identical to those in the Cartesian studies of Tackley (2002), Jellinek and Manga (2002), and Jellinek and Manga (2004). Like the Cartesian studies, the more-dense material accumulated into ridges along the base of the upwelling regions (within the less-dense material) and was swept away from downwellings (within the less-dense material) (**Figure 10(a)–10(c)**). This was an important conclusion because it meant that geometry alone did not play a governing role in the morphology of thermochemical piles. Furthermore, a temperature- and depth-dependent rheology was explored, which included a 30× viscosity increase from the upper to lower mantle and a linearly increasing viscosity with depth in the lower mantle. The main effect of the depth dependence was to increase the size of the thermochemical piles (**Figure 10(d)**). For all of the cases described here, the magnitude of the interface topography was inversely related to the density contrast between layers (larger for smaller density contrasts), and the lateral scale of topographic features was directly related to the size of larger-scale convection cells. The temperature- and composition-dependent rheology cases produced active, dome-like structures described in a later section here.

The numerical 3-D spherical thermochemical convection study of Oldham and Davies (2004) produced similar results, forming thermochemical piles shaped into a network of ridges. They investigated a wide range of initial more-dense layer thicknesses, intrinsic density contrasts between materials, and heating styles. They found that a 1% intrinsic density contrasts led to unstable piles (i.e., the effective density of the more-dense material is less than that of the less-dense material, due to thermal expansion counteracting the intrinsic density). They also found that bottom heating, as opposed to a zero heat flux bottom boundary, required higher intrinsic density contrasts to provide a stable configuration of piles. Intrinsic density contrasts of 4% led to stable configurations in all cases.

The Cartesian and spherical thermochemical studies described in the preceding text were performed in a fluid dynamical context that did not include the influence of plate tectonics. As a result, downwellings and upwellings were relatively evenly spaced, and convection cells had smaller aspect ratios than we think exist in the Earth. McNamara and Zhong

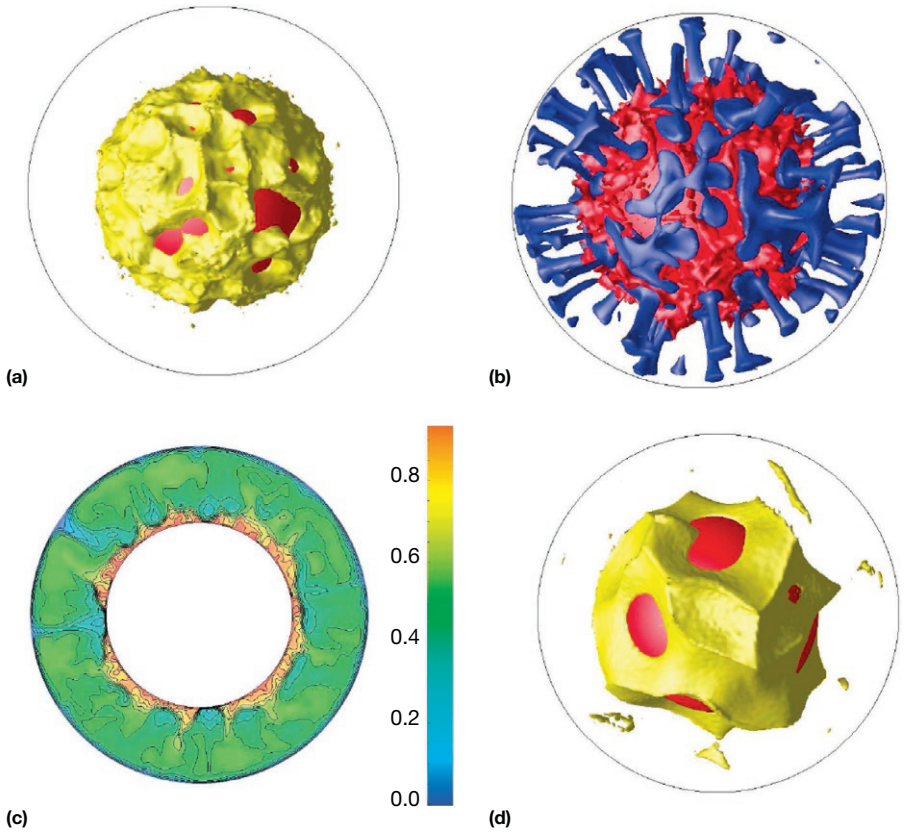


Figure 10 Thermochemical convection calculation in 3-D spherical geometry. The outer black circle represents the equatorial radius (with respect to polar viewpoint). The core is colored red. (a–c) From a calculation that does not have a depth-dependent viscosity. (d) From a calculation that includes a depth-dependent viscosity that includes a viscosity increase into the lower mantle. (a) More-dense material (yellow) accumulates into a ubiquitous network of ridges. (b) Residual temperature isosurfaces that show downwellings (blue) and hotter-than-average material near the CMB (red). (c) Cross section that illustrates the nondimensional temperature field. (d) More-dense material (yellow) for the calculation with depth-dependent viscosity included, in which case the ridges become much larger. Modified from McNamara AK and Zhong S (2004) Thermochemical structures within a spherical mantle: Superplumes or piles? *Journal of Geophysical Research B* 109: B07402, figures 3 and 7.

(2005) hypothesized that convection patterns guided by historical subduction would act to focus the ubiquitous network of ridgelike piles into shapes that better resembled the LLSVPs. They introduced plate history for the past 120 Ma as kinematic velocity boundary conditions to guide the location of subduction, following procedures similar to Bunge et al. (1998) and Lithgow-Bertelloni and Richards (1998). To minimize momentum violation, they had to carefully choose a Rayleigh number that naturally generated convection velocities that were consistent with plate velocities. They found that the more-dense material naturally formed ridgelike morphologies; however, these ridges were focused away from subducting slabs and into regions beneath Africa and the Pacific. Furthermore, the resultant African structure was linear with a NW–SE trend, and the resultant Pacific structure was a superposition of several ridges that combined to form a more rounded structure (**Figure 11**). Mantle plumes (in the less-dense material) were rooted to the tops of the piles, particularly in regions where ridges intersected. These first-order results that position piles beneath Africa and Pacific and produce a linear NW–SE trend to African structure appear robust; however, morphological details, including the position of plumes, depend highly on material properties used, as demonstrated by other studies (e.g., Bower et al., 2013; Garnero and McNamara, 2008; Zhang et al., 2010). In particular, the depth dependence of viscosity appears to play a large role in controlling the size of the ridges that focused into these regions.

Keeping in mind that the detailed morphology of individual thermochemical structures highly depends upon the material parameters used, Ritsema et al. (2007) and Bull et al. (2009) investigated how thermochemical models with plate history as kinematic velocity boundary conditions compare to actual seismic tomography. **Figure 12(a)** and **12(b)** shows thermochemical piles and the temperature field above the CMB from a calculation with parameters similar to those used in McNamara and Zhong (2005). Assuming a pyrolite composition, Bull et al. (2009) converted temperature and pressure to shear-wave velocity using the thermodynamic methods provided by Stixrude and Lithgow-Bertelloni (2005). Seismic tomography provides a blurred image of actual anomalies due to the heterogeneous distribution of earthquake and seismic station locations, so they applied the resolution operator of the tomography model S20RTS (Ritsema et al., 1999, 2004) to convert the geodynamically predicted shear-wave-velocity field to synthetic tomography (**Figure 12(c)**). **Figure 12(d)** compares the power spectrum of the prefiltered and postfiltered

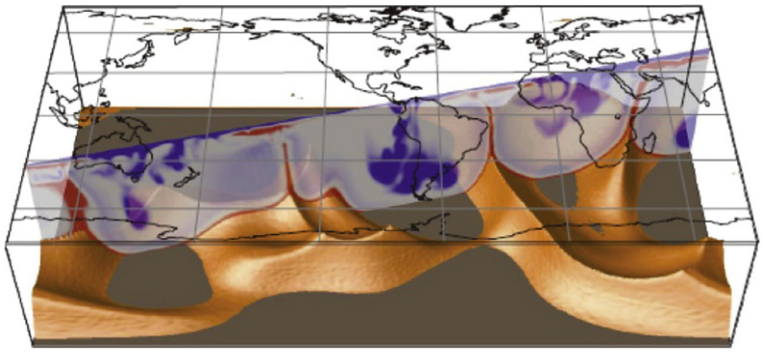


Figure 11 Thermochemical convection calculation in 3-D spherical geometry (unwrapped unto a Cartesian box) that employs 120 My of plate history as kinematic surface velocity boundary conditions. The CMB is colored gray-brown. Thermochemical piles (gold) accumulate into a superposition of ridges beneath Africa and the Pacific. A cross section of the temperature field spans a great-circle plane through the model, illustrating downwellings in subduction regions and thermal plumes that form along pile tops (cold is blue and hot is red). Figure is taken from Garnero and McNamara (2008) and is similar to that performed in McNamara and Zhong (2005).

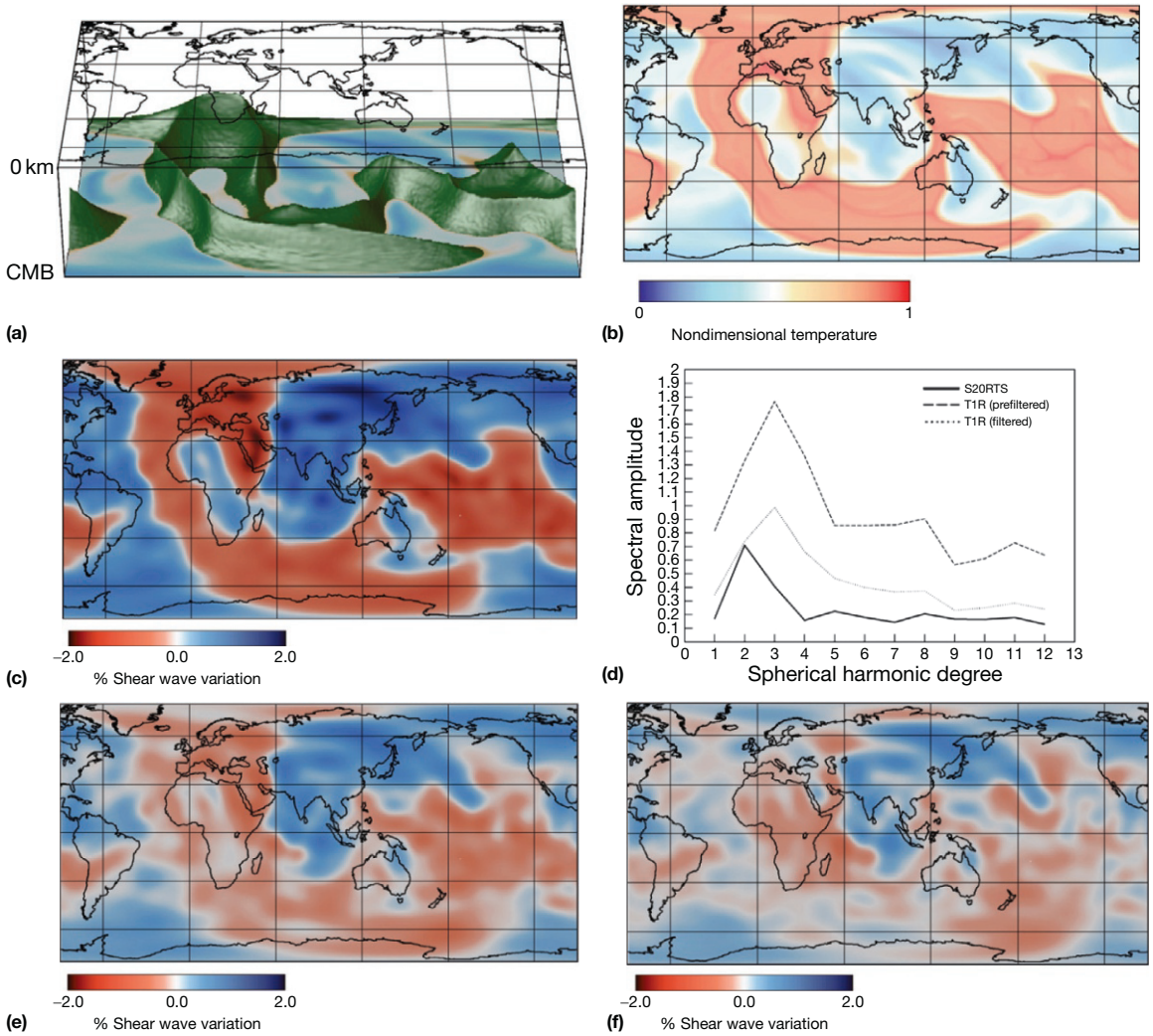


Figure 12 Synthetic tomography created from 3-D spherical geodynamical calculations of thermochemical piles resulting from using 120 My of plate history as kinematic surface velocity boundary conditions. (a) Thermochemical piles (green). (b) Temperature field above the CMB (at 2750 km depth). (c) Synthetic shear-wave tomography produced using the temperature field of the geodynamical calculation. (d) Amplitude spectrum of spherical harmonic decompositions of the tomography model (bold) (Ritsema et al., 1999, 2004), prefiltered temperature field (bold dashed), and synthetic tomography (light dashed). (e, f) Synthetic shear-wave tomography produced using the temperature field of geodynamical calculation and by assuming that thermochemical piles has a 3% and 4% intrinsic increase of shear-wave velocity over background mantle material, respectively. Reproduced from Bull AL, McNamara AK, and Ritsema J (2009) Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters* 278(3–4): 152–162, figures 2, 4, and 5.

geodynamically predicted shear-wave-velocity fields to that of S20RTS. The tomography model is characterized by large amplitude at degrees 2 and 3 followed by a relatively low-amplitude flat spectrum at higher degrees. The geodynamical model also predicted high amplitudes at degrees 2 and 3; however, unlike the tomography model, it exhibited greater amplitude in degree 3 than in degree 2, and the spectrum at higher degrees was not as flat. By extending plate motions back to 450 Ma, similar calculations of Zhang et al. (2010) were able to reproduce a maximum amplitude in degree 2, even without filtering. The main effects of the tomographic filtering were a reduction in the overall amplitudes and a softening of the variability in the higher-degree spectrum. If thermochemical piles are enriched in silicon, by containing a higher fraction of magnesium silicate perovskite, they could have an increased intrinsic shear-wave velocity (e.g., Kellogg et al., 1999). Bull et al. (2009) found that increasing the intrinsic shear-wave velocity caused the thermochemical piles to appear more diffuse in their predicted seismic tomography (**Figure 12(e)** and **12(f)**).

Deschamps and Tackley (2008, 2009) provided an exhaustive set of numerical thermochemical calculation results aimed at documenting pile morphology, stability, and longevity as a function of rheological properties (temperature dependence, compositional dependence, and depth dependence), viscosity increase from upper to lower mantles, intrinsic density contrast between materials, Rayleigh number, internal heating, Clapeyron slope of the 660 km phase transition, and original layer thicknesses. Furthermore, the spectrum of chemical and thermal heterogeneity as a function of depth is compared to probabilistic tomography models (e.g., Trampert et al., 2004) to assess applicability of each to the actual Earth. They found several key observations. Firstly, they found that the intrinsic density contrast between materials played the largest role, and density contrasts in the range of 90–150 kg m^{-3} led to the formation of thermochemical piles (lower contrasts were unstable and higher contrasts led to stratification). A large temperature dependence of viscosity (greater than 10 000× variation) allowed for the formation of long-lived thermochemical piles that also satisfied probabilistic tomography. The same could be achieved without temperature dependence of viscosity if the intrinsic (i.e., compositional) viscosity ratio between the materials was in the range of 0.1–10. The detailed morphology of piles could be significantly affected by compositional viscosity. The original thickness of the more-dense layer had a second-order influence, and the results were relatively insensitive to internal heating. Interestingly, they found that the negative Clapeyron slope associated with the 660 km phase transition played a significant role toward the stability of thermochemical piles, and they suggested that values in the range −3.0 to −1.5 MPa K^{-1} could lead to compositional heterogeneity that would be satisfied by probabilistic tomography. In Deschamps et al. (2011), they further investigated whether long-lived thermochemical pile models that satisfy probabilistic tomography can also satisfy geochemical observations. Citing that geochemical observations of OIBs provide an upper bound on the amount of primitive material included within OIB source material (to be about 10%), they measure the amount of pile material entrained into thermal plumes that form along pile tops. For the cases they investigate, the thermal plumes contain 2–9% pile material by entrainment, satisfying geochemical constraints. However, it should be noted that constraining the amount of entrainment is difficult in convection models due to the high resolution required.

Nakagawa and Tackley (2004a) investigated how thermochemical models affect cooling of the core. They used numerical mantle convection calculations to provide the thermal boundary conditions (temperature and heat flux) on a core heat balance formulation that included core cooling and inner-core solidification. They investigated three kinds of conceptual mantle models: isochemical, primordial thermochemical piles, and accumulation of subducted crustal material. Starting with an entirely liquid core, they performed calculations that spanned the age of the Earth. As constraints for a successful model, they used present-day values of inner-core radius and surface heat flow and the condition that the CMB heat flux must have been large enough to generate a magnetic field for the last 3 Gy. They found that isochemical models lead to a much larger inner core, indicating that they cooled too quickly. However, both types of thermochemical model acted to partially insulate the core and reduce cooling. They found that a 1.5–2% intrinsic density between materials best satisfied the constraints.

Zhang et al. (2010) performed thermochemical calculations that were a significant advancement over those of McNamara and Zhong (2005) because they employed much longer plate history (450 My) and produced structures that more closely resembled seismic tomography. Their calculations showed that if thermochemical structures are governed by plate motions, then the African structure is expected to be significantly younger than the Pacific structure. They showed that prior to Pangea assembly, the African region was characterized by downwelling, and intrinsically dense material accumulated within the Pacific region. Only after Pangea started to break apart did the African structure begin to form. CMB heat flux was explored in Zhang and Zhong (2011). Their models produce a distribution of seafloor age and surface heat flow consistent with present-day observations. They found CMB heat flux to be greatest in paleosubduction regions where downwellings came into contact with the core. In their thermochemical calculations, piles locally reduced CMB heat flux and acted to insulate the core. Also, they find minimums in equatorial CMB heat flux at times that correspond to the Kiaman and Cretaceous superchrons, hinting that reduced CMB heat flux may be related to geodynamo processes.

Lassak et al. (2007, 2010) attempted to use CMB topography as a constraint to test whether the LLSVPs can be caused by thermochemical piles. Earlier work on isochemical convection showed that downwellings press upon the core, causing downwarped topography, and upwelling plumes pull on the core, causing upwarped topography (e.g., Hager and Richards, 1989; King, 1997). It was unclear how thermochemical piles would affect CMB topography, and it was a common speculation in the community that 'dense piles' would cause down-warped topography because they are 'heavy'. In contrast, Lassak et al. (2007) hypothesized that thermochemical piles would instead cause upwarped CMB topography. They reasoned that in order for downwellings to sweep aside pile material, the downwellings must be more negatively buoyant than the pile material. Furthermore, in order for the pile material to develop large interface topography and accumulate into discontinuous piles

along the CMB, their intrinsic density increase must be significantly reduced by thermal expansion, leading to piles having a near-neutral (but slightly negative) buoyancy. Lassak et al. (2007) performed isochemical and thermochemical numerical convection calculations in a 2-D Cartesian geometry. Note that the magnitude of CMB topography is poorly constrained by modeling because it is directly dependent upon viscosity magnitude (which has large uncertainties), so the focus was to investigate the characteristic pattern of topography in map view, with respect to features such as downwellings, piles, and slabs. They found that CMB topography is strongly sensitive to the temperature dependence of viscosity. Like earlier isochemical studies, they found that isochemical convection leads to downwarped CMB topography beneath downwellings and upwarped CMB topography below upwelling plumes. For thermochemical cases, they found that constant (isoviscous) and weak temperature dependence of viscosity (<1000× viscosity variation) led to downwarped CMB topography beneath piles and upwarped CMB topography beneath downwellings. However, for more realistic, stronger temperature dependence of viscosity, CMB topography beneath downwellings was mostly downwarped, and CMB topography beneath thermochemical piles was mostly flat-shaped (laterally) and slightly positive. They also investigated relative buoyancy within and outside of piles and found that downwelling buoyancy is on par with thermochemical pile buoyancy. In Lassak et al. (2010), they expanded their earlier work to investigate 3-D spherical geometry and include self-gravitation (Zhong et al., 2008). They employed plate history as kinematic surface velocity boundary conditions that resulted in either plume clusters (for isochemical cases) or thermochemical piles (for thermochemical cases) beneath Africa and the Pacific (**Figure 13(a)** and **13(b)**). They demonstrated that the plume cluster hypothesis and the thermochemical pile hypothesis

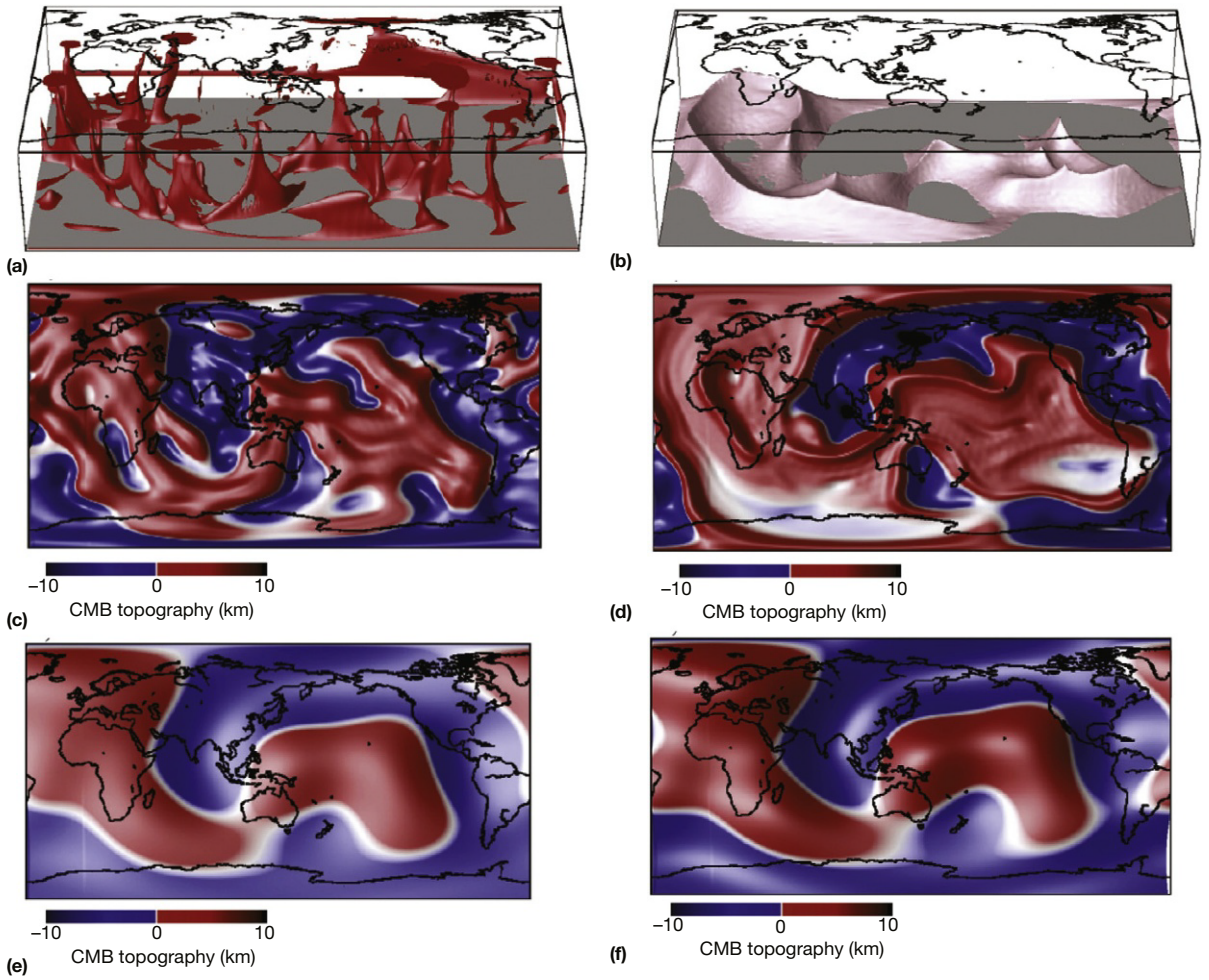


Figure 13 CMB topography predicted for isochemical and thermochemical 3-D spherical convection models. Plate history for the past 120 My was used as kinematic surface velocity boundary conditions. (a) Isosurfaces of higher-than-average temperature, outlining the position of plume clusters. (b) Compositional field illustrating the position of thermochemical piles (pink). (c) CMB topography generated from the isochemical model. (d) CMB topography generated from the thermochemical model. (e) CMB topography generated from the isochemical model, shown only up to degree-4 spherical harmonics. (f) CMB topography generated from thermochemical model, shown only up to degree-4 spherical harmonics. Blue represents downwarped CMB topography and red represents upwarped CMB topography in all cases. Reproduced from Lassak TM, McNamara AK, Garnero EJ, and Zhong SJ (2010) Core–mantle boundary topography as a possible constraint on lower mantle chemistry and dynamics. *Earth and Planetary Science Letters* 289(1–2): 232–241, figures 1, 2, and 5.

each resulted in characteristically different patterns of CMB topography (**Figure 13(c)** and **13(d)**). In both isochemical and thermochemical cases, CMB topography was downwarped beneath downwellings. In the isochemical cases, CMB topography was generally upwarped in plume cluster regions, with a rough relief pattern corresponding to individual plumes (**Figure 13(c)**). In contrast, the thermochemical cases resulted in relatively flat and slightly upwarped CMB topography beneath piles. In addition, thin ridges of higher topography formed along the perimeters of the piles (**Figure 13(d)**). This work showed that if LLSVPs are caused by thermochemical piles, CMB topography should form plateau-shaped upwarped topography beneath them, surrounded by a thin rim of locally elevated relief. Seismic observations of CMB topography are not in great agreement between studies, have highly heterogeneous coverage, and are typically represented in low-degree spherical harmonic projections (e.g., Sze and van der Hilst, 2003). Lassak et al. (2010) decomposed and projected their CMB topography results into only low-order spherical harmonics, similar to seismic studies, and found that although plume clusters and thermochemical piles each produce a characteristic pattern of topography, they were virtually identical at long wavelength (**Figure 13(e)** and **13(f)**). Future seismic studies should target the margins of LLSVPs to search for the narrow rim of elevated topography; confirmation of which would support the existence of thermochemical piles.

Yoshida (2008) performed a set of global, instantaneous flow calculations using tomographically derived density distributions to investigate CMB topography. He explored the addition of lateral viscosity variations, the presence of dense thermochemical piles in LLSVP regions, and a viscosity reduction in the lowermost mantle (i.e., representing a D'' layer). He found that each of these additions acted to reduce the amplitude of CMB topography, and one or more of them are required to satisfy seismological constraints of less than 2 km variation.

McNamara et al. (2010) explored whether ultralow-velocity zones (ULVZs) could be used to test whether thermochemical piles exist within LLSVP regions. Under the assumption that the ULVZ is caused by a small volume of high-density compositional heterogeneity, they performed high-resolution, three-component thermochemical calculations in 2-D Cartesian geometry that contained the (a) background mantle, (b) thermochemical piles, and (c) a thin, 5 km thick layer of high-density ULVZ material (**Figure 14**). They first performed two-component (thermochemical piles within a less-dense, background mantle) calculations until they reached a statistical thermal steady state. They then inserted a 5 km thick layer of ULVZ to investigate where the material would accumulate and what morphology it would develop. The layer of ULVZ material quickly formed multiple accumulations that migrated to the bottom side margins of thermochemical piles due to laterally convergent flow in these locations (e.g., Hernlund and Tackley, 2007). The height and width of accumulated ULVZ largely depend upon density contrast and original volume of material. For the cases studied, typical ULVZ height ranged from 30 to 45 km and ULVZ width ranged from 550 to 1150 km. The calculations were run for the equivalent of several billion years, and during that time, the larger thermochemical piles periodically underwent separations and mergers. If two piles merged, ULVZ material along their margins would be temporarily orphaned within the middle of the new, larger resultant pile before eventually migrating to the margins. Most importantly, it was found that ULVZ accumulations had an asymmetrical thickness, being thicker on the outboard side of the pile and thinner on the inboard side. This is due to differential viscous stress acting on the ULVZ accumulation. Piles are significantly hotter than the background mantle, and temperature-dependent viscosity leads to piles having a much lower viscosity than background mantle. As a result, the inboard pile side of the ULVZ accumulation has a lower viscous stress acting to prop it up than the outboard background mantle side. Future seismic studies should focus on ULVZ shape, and if they observe such an asymmetrical thickness to ULVZ, it would provide strength to the hypothesis that thermochemical piles exist and cause the LLSVPs.

7.11.4.2.4.3 Passive thermochemical piles caused by accumulation of oceanic crust

Another thermochemical hypothesis for the cause of LLSVPs is that they are caused by the accumulation of oceanic crust. Mineral physics experiments indicate that oceanic basaltic crust is expected to become more dense than background mantle at lower-mantle pressures (e.g., Hirose et al., 1999, 2005).

The idea behind this hypothesis is based on the following:

1. Oceanic lithosphere with basaltic crust descends into the lowermost mantle.
2. The oceanic basaltic crust becomes separated from the oceanic lithosphere.
3. The segregated basaltic crust travels to and accumulates at the base of mantle upwelling regions at a rate equal to or greater than the rate at which it is entrained away via mantle plumes.

Geodynamical modeling of this hypothesis is challenging for several reasons. Firstly, the oceanic crust is thin (6 km), which requires high grid resolutions to fully resolve it, particularly as it segregates from the lithosphere. Secondly, with a few exceptions that require special technical innovation (e.g., Crameri et al., 2012), geodynamical models do not self-consistently produce the one-sided, asymmetrical subduction such that occurs on the Earth (see **Chapters 7.07** and **7.09**, this volume). Instead, geodynamical models typically result in symmetrical, double-sided subduction that leads to a sandwiched layer of oceanic crust between two high-viscosity layers of the lithosphere. It is unclear how the high-viscosity lithosphere envelope surrounding the crust affects its ability to segregate at the base of the mantle. Furthermore, the rheology and strength of subducting slabs in the lower mantle are not well understood, and it is unclear how slabs deform in the lowermost mantle (e.g., buckling and pure shear deformation), and the manner of slab deformation will affect the ability, style, and amount of crustal segregation.

Xie and Tackley (2004a,b) performed 2-D thermochemical calculations that introduced basaltic crust and its complementary harzburgite residuum by melt-induced differentiation to investigate chemical evolution (e.g., U–Pb, Sm–Nd, and He systems). As material exceeded the melting temperature, crust and residue were generated by calculating the amount of melt and placing an equivalent amount of basalt at the surface of the

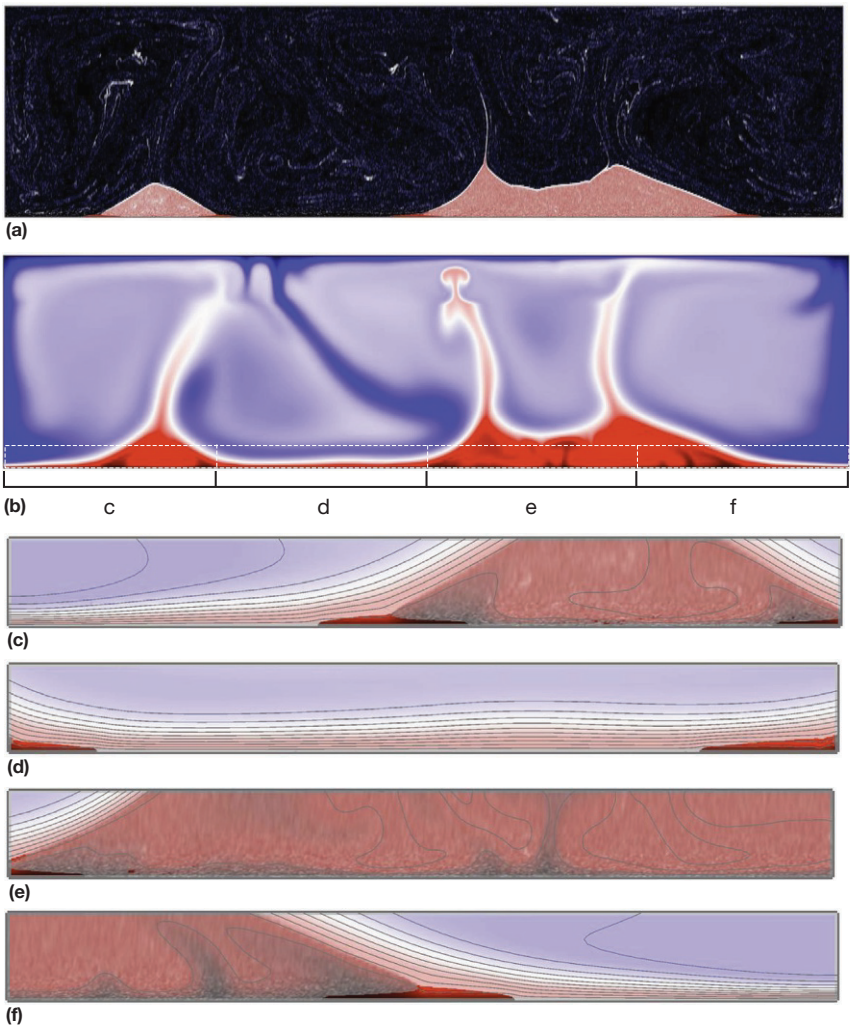


Figure 14 Three-component thermochemical calculation including background mantle, thermochemical piles, and a small volume of high-density ULVZ material. In addition to temperature-dependent viscosity, ULVZ material has an additional 10x intrinsic reduction in viscosity. (a) Compositional field showing background mantle (black), thermochemical piles (red), and ULVZ material (dark red). ULVZ material has accumulated along the CMB at pile margins. (b) Temperature field (red is hot and blue is cold). White boxes at the bottom illustrate the location of zoom panels. (c–f) Zoomed-in panels of the lowermost mantle, illustrating both temperature and compositional fields. ULVZ material accumulates along the margins of thermochemical piles and maintains an asymmetrical shape, being thinner on the side that is in contact with the piles. Reproduced from McNamara AK, Garnero EJ, and Rost S (2010) Tracking deep mantle reservoirs with ultra-low velocity zones. *Earth and Planetary Science Letters* 299 (1–2): 1–9, figure 6.

model. They applied a yield-stress approach to generate subducting slabs (e.g., Tackley, 2000b,c). Starting with a uniform mantle, they found that significant accumulations of more-dense crust could accumulate in the lowermost mantle over geologic time and provide a storage reservoir for heat-producing elements. Characteristic of calculations that involve accumulation of oceanic crust, piles appear quite 'messy,' compared with primordial piles examined in the earlier section, and significant amounts of chemical advection occur across the top interface of piles. This difference is most easily compared in Nakagawa and Tackley (2004b), which examined both primordial piles and accumulation of oceanic crust created by melt-induced differentiation.

Nakagawa and Tackley (2005) performed similar calculations that included the postperovskite phase transition in the lowermost mantle. They found that accumulations of oceanic crust became much hotter than background mantle, likely too hot for the postperovskite phase to occur. Instead, postperovskite mostly occurred at the base of downwelling regions, away from the hot piles. Furthermore, the presence of postperovskite tends to destabilize the accumulations of crustal material because it modifies the lateral density variation at the base of the mantle. Therefore, the presence of postperovskite in the lower mantle requires higher-density contrasts between the basaltic crust and background mantle to form stable, long-lived accumulations.

Nakagawa and Tackley (2006) produced the first 3-D thermochemical calculations involving melt-induced differentiation to generate oceanic crust, in a Cartesian geometry. An important conclusion of this work is that the resultant

accumulations of oceanic crust in their models developed a morphology similar to that in primordial piles, consisting of a network of ridges. Therefore, if seismology better defines the horizontal platform of compositional structures in the lowermost mantle, this cannot be used to discriminate between primordial piles and piles caused by the accumulation of oceanic crust. However, the accumulations of oceanic crust have 'messy' tops compared to primordial piles, so perhaps a careful mapping of LLSVP tops could help differentiate between these scenarios. Nakagawa et al. (2010) greatly expanded this work by employing a 3-D spherical geometry and using self-consistent thermodynamic formulation of mineralogy for a given compositional model (represented by various oxide ratios). They found that the amount of crustal segregation and accumulation was highly sensitive to the compositional model used, each of which provided a different basalt/harzburgite density contrast.

Nakagawa and Tackley (2011) provided an important dynamical observation regarding the role of postperovskite in the amount of crustal segregation. Motivated by mineral physics experiments that indicated that postperovskite may be weaker than perovskite (e.g., Ammann et al., 2010; Hunt et al., 2012), they introduced the postperovskite phase transition and prescribed that phase a lower intrinsic viscosity (1000× lower). As shown in their earlier work, the transition primarily occurred at the base of downwelling regions. This weakening allowed the crust to more easily segregate from the downwelling slab, producing a larger accumulation of crust in the lowermost mantle. The importance of postperovskite weakening was confirmed in the later study of Li and McNamara (2013). Weakening of the slab allows crust to more easily penetrate into the lowermost thermal boundary layer, where it can more easily survive against viscous stirring.

Brandenburg and van Keken (2007a,b) developed 2-D cylindrical models that employed a force-balanced method of plate generation at the surface. Their model provided a greater control of oceanic crust generation and thickness by prescribing a 10 km thick crust (and harzburgite below) within discrete zones located where plates diverge. They recognized an important modeling consideration when doing this kind of work. Model boundaries (or internal symmetry condition in their case) can artificially increase the amount of crustal accumulation, for reasons likely similar to those that allow weakened postperovskite to facilitate more accumulation. The crust that descends with a downwelling along a vertical model boundary is protected from viscous stirring because it is guided, undisturbed, along that boundary. When it arrives in the lowermost mantle, it is directly inserted into the lowermost thermal boundary layer and is therefore not easily stirred by viscous stresses. This was also observed in the preliminary work of Li and McNamara (2013), which was ultimately rectified by removing crustal tracers that entered a thin buffer zone along the vertical boundaries. Brandenburg and van Keken (2007a) found that large accumulations of oceanic crust could form at Earth-like, high Rayleigh numbers only if the intrinsic density between the crust and background mantle was larger than that inferred from mineral physics experiments. Brandenburg et al. (2008) provided a nice culmination of their work, in which they investigated the evolution of several isotopic systems and found that accumulations of crust in the lowermost mantle provide isotopic signatures comparable to the HIMU and EM-1 geochemical components observed in OIBs.

Davies (2008) demonstrated that depth-dependent density contrast between the subducted oceanic crust and background mantle may have led to mantle layering in the early mantle, when it was hotter. Mineral physics experiments indicate that while oceanic crust is expected to be more dense than the background mantle in the upper mantle and most of the lower mantle, it may become less-dense directly below the transition zone, between 660 and 750 km depth. His calculations demonstrated that in an earlier, hotter mantle, this density inversion could cause subducted basalt to be trapped, by what he dubbed a 'basalt barrier mechanism.' The resultant layered convection would lead to a strong temperature contrast between the upper and lower mantles, and layering would episodically break down and avalanching would occur. Eventually, as the mantle cooled, the subducting lithosphere became negatively buoyant enough to penetrate this barrier, and whole mantle convection was established. This transition occurred at about 1.6–1.8 By in his models. An important consideration of this work is that even if oceanic crust is too thin to accumulate in the lower mantle in a steady-state process that is occurring today (e.g., Li and McNamara, 2013), this type of mechanism could facilitate lower-mantle accumulation in the past by flushing large volumes of trapped upper-mantle basalt into the lower mantle.

Tackley (2011) performed 2-D and 3-D thermochemical calculations in which a slab segment (including both basalt and harzburgite) is placed into the mid-mantle and is allowed to descend to the CMB. Under the assumption that slabs thicken as they enter the lower mantle due to a viscosity increase, motivated by Gurnis and Hager (1988), he employed a crustal thickness of 30 km, with a 300 km thick harzburgite layer beneath. The slab segment was inserted at various angles, ranging from shallowly dipping to steeply dipping. He found that the density variation within the slab (dense basalt overlying less-dense harzburgite) caused a torque-induced rotation. For steeply and moderately dipping slabs, this rotation acted to flip the slap upside down such that basalt was on the bottom and harzburgite was on the top. Upon heating at the CMB, the basalt would segregate from the slab, and plumes of harzburgitic material would form plume heads that would entrain some of the basaltic material in the plume tails. For more shallowly dipping slab segments, the slabs landed at the CMB with the basalt side up. In these cases, plumes of harzburgite would escape through the edges of the slab segment and/or form plumes that burst through the basaltic layer.

Li and McNamara (2013) examined whether realistically thin (6 km) oceanic crust could accumulate into large, LLSVP-sized structures within upwelling regions of the lowermost mantle. They performed high-resolution 2-D Cartesian convection calculations, continually introducing a more-dense oceanic crust in the upper 6 km of the model, which was subducted into the lower mantle by downwellings. They explored several buoyancy numbers for the crust, leading to density contrasts ranging from 0% to 3%. Similar to previous studies (e.g., Nakagawa and Tackley, 2011), they found that weakened postperovskite significantly increases the amount of crustal segregation that can occur. In all cases studied, they found that viscous forces exceeded the negative buoyancy

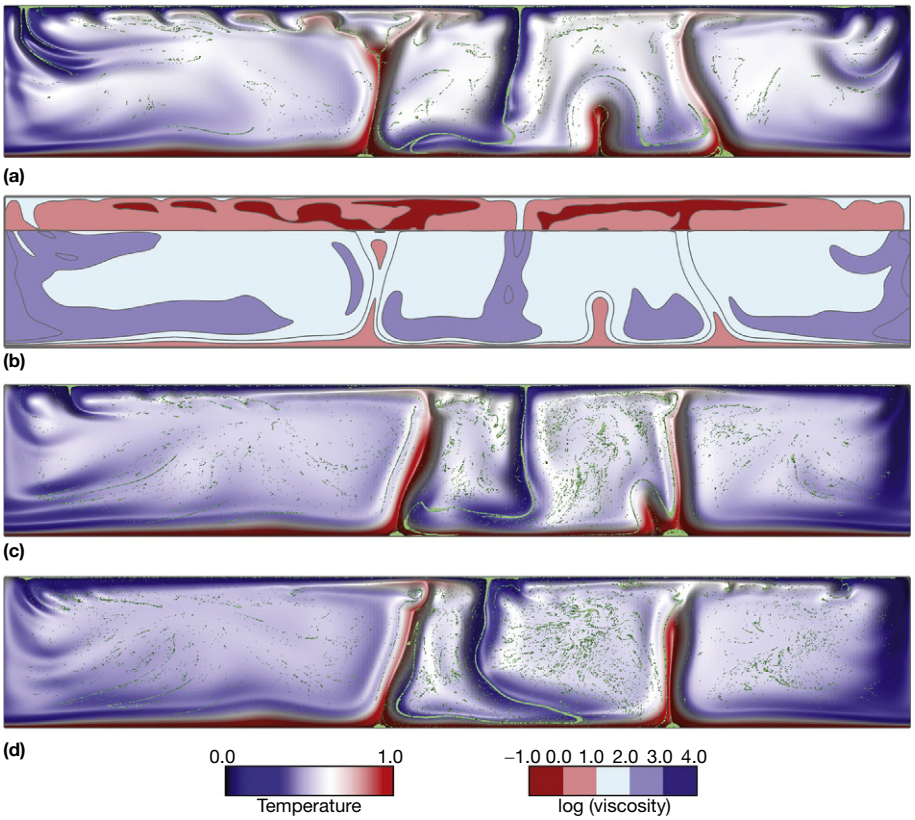


Figure 15 Thermochemical calculation that includes background mantle and a thin, 6 km thick more-dense oceanic crust generated at the surface. (a) Nondimensional temperature field with the oceanic crust (shown in green) superimposed, shown at 1.0 Gy from the start of the calculation. (b) Logarithm of the nondimensional viscosity, shown at 1.0 Gy from the start of the calculation. (c) Nondimensional temperature field with the oceanic crust (shown in green) superimposed, shown at 2.0 Gy from the start of the calculation. (d) Nondimensional temperature field with the oceanic crust (shown in green) superimposed, shown at 2.8 Gy from the start of the calculation. Modified from Li M and McNamara AK (2013) The difficulty for subducted oceanic crust to accumulate at the Earth's core–mantle boundary. *Journal of Geophysical Research*, *Solid Earth* 1–10, figure 1.

forces associated with the more-dense crust and the majority of oceanic crust was easily stirred into the background mantle. A small accumulation (<100 km thick) of crustal material did accumulate at the base of plumes, but the volume did not increase over geologic time due to entrainment into the plumes (**Figure 15**). In conclusion, they found that 6 km thick oceanic crust was too thin to accumulate into large structures on the same spatial scales as LLSVPs. Therefore, if LLSVPs are caused by an accumulation of oceanic crust, they must have formed by thicker accumulations of crust entering the lowermost mantle, which could occur if the oceanic crust was significantly thicker in earlier times and/or if the oceanic crust accumulated in the transition zone before being flushed into the lower mantle (e.g., Davies, 2008).

7.11.4.2.4.4 Combined models: Primordial piles and subducted oceanic crust

If LLSVPs are caused by the accumulation of intrinsically more-dense material, the most general scenario is that they were formed by several processes (e.g., Tackley, 2012), including early differentiation processes and events along with the accumulation of thickened, ancient oceanic crust. Furthermore, the oceanic crust is continually being introduced into the lower mantle through subduction. Therefore, a more appropriate thermochemical model of the Earth's mantle probably should include three components: (1) background mantle that provides the MORB source; (2) an ancient, more-primitive (loosely defined) accumulation of more-dense material; and (3) geologically young oceanic crust being introduced at downwellings (e.g., Tackley, 2000a).

Samuel and Farnetani (2003) and Farnetani and Samuel (2003) were perhaps the first to investigate such three-component thermochemical systems. In Samuel and Farnetani (2003), they performed 2-D Cartesian thermochemical models that included a more-dense basal layer, background mantle, and subducted oceanic lithosphere (that included both the more-dense crust and subcrustal lithosphere without an intrinsic density increase). Their basal layer had a volume equivalent to 25% of mantle volume, and although downwellings could locally depress it, this layer appeared to remain largely continuous throughout the lowermost mantle, as opposed to distinct, isolated piles. Mantle plumes formed along the top of this layer, entraining more-dense material into them. Compositional advection was performed by Lagrangian tracers that also stored the time evolution of U, Th, K, He-3, and He-4. By examining the He ratios in the uppermost mantle beneath divergent regions and within mantle plumes, they constructed histograms that demonstrated that such a

conceptual model predicts that MORBs maintain a narrow range of He-3/He-4 values while OIBs exhibit highly variable ratios, both of which are consistent with geochemical observations. Furthermore, they demonstrated that mantle plumes are expected to entrain multiple compositional components: the more-dense material (i.e., an undegassed, primitive reservoir), oceanic crust, and oceanic lithosphere (i.e., harzburgite).

Nakagawa and Tackley (2004b) employed 2-D cylindrical geometry models, and in their cases, the density of the oceanic crust is significantly higher than the primitive material, both of which are higher than the background mantle. They based the buoyancy number for the primitive material to coincide with estimated density differences from seismic tomography (Forte and Mitrovica, 2001), and they made the buoyancy number for the crustal material large enough to form piles of oceanic crust over geologic time. The primitive material was not dense enough to remain stable, and it formed blobs throughout the mantle, and the crustal material was swept into piles in the lowermost mantle. They found that their resultant compositional structures resembled the depth distribution of seismic heterogeneity (from seismic tomography) and provided a multireservoir model that could satisfy geochemical observations.

Tackley (2011) performed 2-D cylindrical and 3-D spherical calculations of basalt segregation in the presence of a preexisting dense layer residing in the lower 150 km of the mantle (continuation of the work described in the previous section). The initially uniform-thickness dense layer was swept away from the descending slab that included a 30 km thick more-dense oceanic crust. Interestingly, he found that the amount of the lowermost mantle crustal segregation increased as a result of including the preexisting dense layer (compared with identical cases without the preexisting dense layer). In these cases, a fraction of the basalt ended up residing as small subpiles within the preexisting layer. Furthermore, he found that 2-D versus 3-D geometry plays a strong role in the fraction of basalt that segregated from the slab. In 3-D geometry, 50–70% of the crust segregates from the slab in the presence of a preexisting dense layer, compared with only 25–45% in the absence of a preexisting layer. In 2-D, the amount of basalt segregation increases in the presence of a preexisting layer and decreases in its absence.

Li et al. (2014) investigated the evolution of such a three-component thermochemical system over geologic timescales in high-resolution 2-D Cartesian geometry. To generate an initial condition, they first performed a two-component thermochemical calculation that included more-dense material (representing a more-primitive reservoir, loosely defined) and background mantle (representing MORB source) until the system reached a statistically steady thermal state. They employed a buoyancy number ($B = 0.8$) that leads to the formation of distinct, thermochemical piles with thermal plumes rooted along their top cusps. From that initial condition, they performed a series of calculations in which the more-dense oceanic crust was continually introduced into the upper 6 km of the model, which would get transported to the lower mantle via downwellings. The model also included a weakening of postperovskite (which occurred at the base of downwellings), motivated by previous studies that demonstrated its strong role in crustal segregation (e.g., Li and McNamara, 2013; Nakagawa and Tackley, 2011). They varied (1) the density of oceanic crust (ranging from less-dense to more-dense than the more-primitive, thermochemical piles) and (2) the viscosity reduction associated with postperovskite regions. **Figure 16** is a snapshot of their reference case, in which the oceanic crust and more-primitive pile material have the same intrinsic density. In all cases they investigated, they found similar fundamental dynamics characterized by the subducted oceanic crust following one of several pathways in the lower mantle. Oceanic crust that did not reach the lowermost thermal boundary layer right above the CMB did not efficiently segregate from the downwelling slab, and as a result, it was largely stirred directly into the background mantle, forming a 'marble-cake'

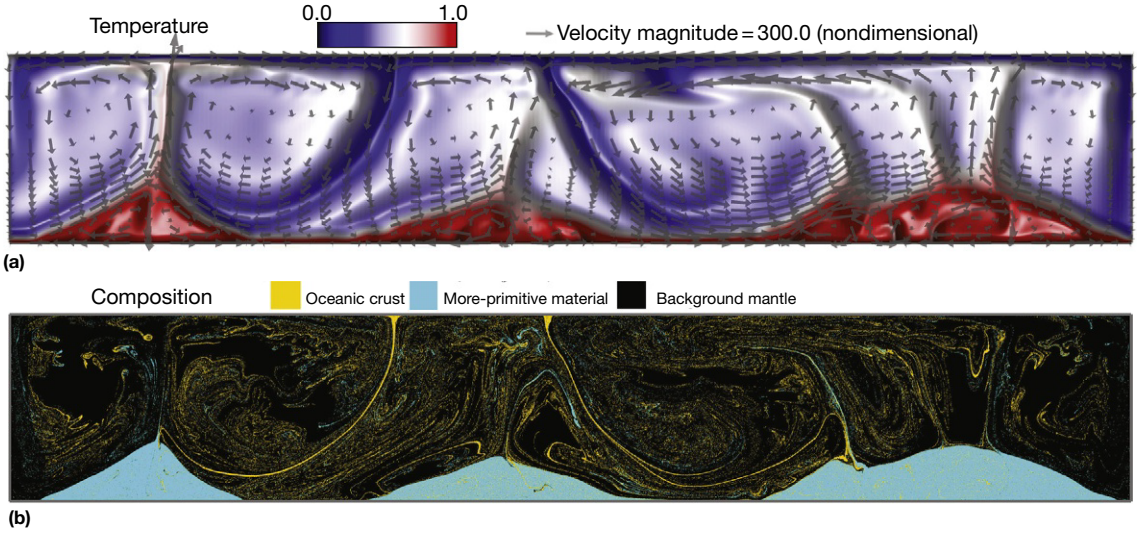


Figure 16 Three-component thermochemical calculation illustrating the interaction of subducted oceanic crust with primordial piles. (a) Nondimensional temperature field with the vectors representing nondimensional viscosity superimposed. (b) Compositional field showing background mantle (black), primordial piles (cyan), and oceanic crust (yellow). Modified from Li M, McNamara AK, and Garnero EJ (2014) Chemical complexity of hotspots caused by cycling oceanic crust through mantle reservoirs. *Nature Geoscience*, http://dx.doi.org/10.1038/ngeo2120.

compositional texture. Oceanic crust that did reach the lowermost thermal boundary layer was then transported laterally toward and up along the top boundaries of the thermochemical piles. Once the oceanic crust reached the top cusp of a thermochemical pile, where thermal plumes are rooted and entrain material from the pile, some fraction of it was entrained directly into the plume. The remainder accumulated at the top of the pile (at the base of the plume) until it was flushed into the pile. Once inside the piles, the oceanic crust was progressively stirred, becoming stretched and thinned with time. Some of this crust would ultimately be entrained, along with the more-primitive pile material, out of the pile and into thermal plumes. This characteristic behavior occurred over the lifetime of the calculations (representing several billion years of geologic time), and as a result, the composition of both thermochemical piles and the background mantle becomes more enriched in the oceanic crust. Assuming that primordial thermochemical piles cause the LLSVPs, this work resulted in the following conclusions:

1. The oceanic crust subducted into the lower mantle is either (a) directly stirred into the background mantle, (b) entrained directly into mantle plumes, or (c) flushed into thermochemical piles (and could be later entrained out into mantle plumes).
2. Mantle plumes are expected to include the following in variable amounts: (a) background mantle, (b) geologically young oceanic crust (which is entrained directly into mantle plumes), and entrained pile material that includes (c) more-primitive material and (d) ancient oceanic crust (that had been flushed into the piles at an earlier time). For the cases studied, mantle plumes contained about 2% more-primitive material and approximately 3% oceanic crust; however, it is important to note that these are likely upper bounds because it is difficult to constrain entrainment in numerical models, even those with high resolution.
3. The fraction of the oceanic crust that gets flushed into piles increases as (a) the density contrast between oceanic crust and more-primitive material increases and (b) the viscosity of postperovskite becomes smaller.

7.11.4.2.4.5 Active (or metastable) thermochemical superplumes

The previous sections highlighted thermochemical scenarios in which the accumulations of more-dense material behaved as long-lived and radially stable, passive structures that are easily swept away from downwellings into upwelling regions. The primary cause for this passive behavior seems to be a combination of their persistent negative buoyancy and weakening due to temperature dependence of viscosity. This section highlights thermochemical work on active structures, which have positive buoyancy and/or a rheology that leads to their having higher viscosity than the surrounding mantle. These types of structures have a rich variability in their morphology and behavior.

Davaille and colleagues had arguably performed the most extensive amount of geodynamical work on active thermochemical structures, performing laboratory tank experiments (e.g., Cadio et al., 2011; Davaille, 1999; Davaille et al., 2002, 2003, 2005; Le Bars and Davaille, 2002, 2004a,b). This collection of work well documents the behavior of both passive and active thermochemical structures as a function of dense layer thickness, viscosity, and intrinsic density. Much of this work is summarized in **Chapter 7.03** of this volume and will not be discussed in detail here. One of the more interesting aspects of that work is that it demonstrates the feasibility that LLSVPs could be caused by thermochemical structures that undergo alternating positive/negative buoyancy through time, resulting in cycles of rising and sinking superplumes. While being intrinsically more dense than the background mantle, as this material heats up at the CMB, thermal expansion causes it to have a net positive buoyancy, forming actively rising superplumes of material. As this material reaches the upper regions of the mantle, it cools and becomes more dense and sinks back down again with negative buoyancy. They demonstrate that this rising/sinking process can occur several times over geologic history before the structures are fully stirred into the mantle and that these structures may explain seismic tomography images (e.g., Davaille et al., 2005). The reader is referred to **Chapter 7.03** of this volume for a more comprehensive description.

McNamara and Zhong (2004) investigated thermochemical convection using both temperature-dependent and composition-dependent viscosities within a 3-D spherical geometry. They performed cases in which the more-dense material had a 100× or 500× intrinsic viscosity increase with respect to the surrounding, background mantle. Combined with temperature-dependent viscosity, this produced unique thermochemical structures consisting of weak interiors and strong 'rinds'. Because the more-dense material maintained a high temperature, the temperature dependence of viscosity had a stronger effect than the compositional dependence within the piles themselves. However, along the margins of the structures, which are cooler than the interior, the compositional dependence dominated, forming the stronger 'rind'. As a result, the more-dense material formed long-lived, rounded domes that maintained a relatively constant height and did not oscillate like the domes of Davaille (1999); however, they migrated laterally around the lower mantle. These domes contained single convection cells, with the upwelling portion located at the center of the dome.

Using seismic tomography to provide density structure, Forte and Mitrovica (2001) computed global flow models that employed radial viscosity profiles (without temperature dependence) constrained by gravity, plate motions, surface topography, and CMB ellipticity. Their results indicated that thermal buoyancy exceeded the negative compositional buoyancy of LLSVP regions and that these regions were actively rising through the mantle. Simmons et al. (2007) advanced this approach, and focusing on the African LLSVP, they concluded that it is best explained as a positively buoyant superplume rising from a negatively buoyant pile residing at the base of the mantle.

Ni et al. (2002) performed 2-D cylindrical thermochemical calculations to investigate the cause of the possible vertical, eastward tilt associated with the African LLSVP, as observed in seismic tomography. They found that conventional Boussinesq calculations failed to produce a tilted thermochemical structure with roughly parallel sides, in the presence of imposed surface plate velocity, unless it had a relatively low buoyancy

number resulting in the structure being unstable. This would imply that if the African LLSVP is caused by a thermochemical pile, it is fortuitously caught in act of overturn, transitioning into an unstable structure. While not impossible, this conclusion is not entirely satisfactory because it implies that we are currently observing a coincidental, transitory event in the African LLSVP. Tan and Gurnis (2005) (concurrently with Samuel and Bercovici, 2006) found a solution to this problem by performing compressible convection calculations in which the more-dense material had a higher bulk modulus than the background mantle, consistent with seismological observations. The higher bulk modulus in the pile material makes it less compressible than the surrounding, background mantle, resulting in a depth-dependent density contrast between the two materials that becomes smaller with depth. This allows for combinations of parameters that produce thermochemical structures that have a height of neutral buoyancy within them, below which pile material becomes less dense than the background mantle and above which pile material becomes more dense than background mantle. The resulting structures are effectively stable overall but have properties of unstable thermochemical structures, particularly in that they have high-standing vertical sides, consistent with seismic tomography observations of the African LLSVP. They refer to these structures as 'metastable superplumes.' Tan and Gurnis (2007) provided an extensive continuation and parameter exploration of this work. In Sun et al. (2007a), they found that such structures can explain various seismic properties of the African LLSVP such as the anticorrelation between bulk and shear-wave velocities and SKS travel times.

Tan et al. (2011) further explored metastable superplumes in 3-D spherical geometry, employing compressibility. As in their earlier 2-D work, the thermochemical structures formed steep-sided dome-like structures. They found that the structures resided away from subduction regions and could survive for billions of years. Furthermore, mantle plume preferentially formed along the margins of the metastable superplume structures, unlike conventional primordial thermochemical piles that typically have plumes rooted along their tops. This is a significant finding given that hot spot locations appear to be statistically situated more along the margins of the LLSVPs than their interiors (e.g., Burke and Torsvik, 2004; Thorne et al., 2004). They also found that metastable superplumes could remain relatively stationary (on a timescale of hundreds of millions of years) if they are located away from subduction; however, changing subduction patterns could cause them to quickly move laterally to maintain their distance from downwellings.

Bower et al. (2013) investigated metastable superplumes in 3-D spherical geometry that employed plate motions for the past 250 million years as kinematic surface velocity boundary conditions. Their plate velocities were distributed into 1 million year intervals, allowing a more continuous and natural development of downwelling structures. This is a significant advance over previous methods (e.g., McNamara and Zhong, 2005) that divided plate motion history into a small number of plate motion configuration stages, each of which lasting for long periods of geologic time and, as a result, generating discontinuous downwelling structures. Furthermore, Bower et al. (2013) recognized that they could generate metastable superplumes using the Boussinesq approximation (which, by definition, neglects compressibility). They achieved this by directly assigning the depth-dependent density contrast between the more-dense material and background mantle that would have self-consistently occurred due to the two materials having different bulk moduli in compressible convection. This is important in two regards. Firstly, it demonstrates that morphological properties of metastable superplumes are physically governed by the depth-dependent density contrast caused by compressibility rather than compressibility itself. Secondly, it demonstrates that such structures can be generated in Boussinesq convection codes. Their study found that plate motions guided the location of metastable superplumes into a ridgelike structure beneath Africa and a rounded structure beneath the Pacific, with plumes forming along the margins of these structures.

7.11.4.3 Long-Term Stability of LLSVPs

Thorne et al. (2004) noted the geographic correlation between the surface locations of the present-day hot spots and the margins of the LLSVPs (defined by the strong gradients in shear-wave tomography that surround the LLSVPs). For hot spots overlying LLSVP regions, they found that hot spots were more than twice as likely to overlie the margins. If one assumes that hot spots are caused by nearly vertical mantle plumes originating in the lowermost mantle, this correlation hints at a physical connection between LLSVP margins and hot spots. In addition to noting this for present-day hot spot locations, Burke and Torsvik (2004), Burke et al. (2008), Torsvik et al. (2008), and Torsvik et al. (2010) found a statistical correlation between the original eruption locations (based on paleomagnetic reconstruction) of LIPs and kimberlites with the margins of the present-day LLSVP locations. The LIP and kimberlite ages extend back to 500 Ma, and it is intriguing that these volcanic features may have originated from these specific locations throughout the Phanerozoic. Assuming a physical connection between these volcanic features and LLSVP margins (based on present-day hot spots), these correlations hint at the possibility that LLSVPs may have remained in the same locations throughout this time.

Conrad et al. (2013) used plate motions since 240 Ma to infer large-scale properties global mantle flow patterns throughout this time. They deconstructed present-day plate velocities into dipole and quadrupole components, which are related to degree-1 and degree-2 convergence and divergence. They found that the African and Pacific regions overlying the LLSVPs are characterized by quadrupole divergence. They perform a similar analysis on the driving forces associated with slab pull and basal tractions on plates, derived from a geodynamical model that uses seismic tomography to infer the density field. There is a general agreement between the dipole and quadrupole locations between the present-day plate motions and the driving forces generated from the geodynamical model, particularly for the quadrupole convergence beneath western Pacific and South America and divergences above the LLSVPs. This agreement provides evidence that plate motions and mantle flow are linked, at least for the present day. They continued this analysis for plate reconstructions extending over the past 240 million years. They found that the

quadrupole divergences remain over the Pacific and Africa throughout, and the quadrupole divergences tend to wander about the margins of the Paleo-Pacific. They used this result to conclude that upwelling has remained stable over the past 240 million years beneath Africa and the Pacific. Furthermore, they speculated that whatever is causing the LLSVPs may be 'stable anchors' that organize mantle convection and maintain upwelling above these regions.

Given our current understanding of thermochemical convection, it is difficult to imagine how thermochemical structures could form 'stable anchors.' From a fluid dynamical standpoint, thermochemical structures appear to be guided by changing subduction patterns, even for active structures. As described earlier, Zhang et al. (2010) carried out spherical thermochemical calculations using 450 My of plate history as surface kinematic boundary conditions. Their calculations showed that while plate motions predict a relatively stationary location for the Pacific structure over that time, it is difficult to reconcile an equally old age for the African structure. Instead, the African structure was created after the breakup of Pangea in their calculations.

7.11.4.4 Summary and Future Work

Over the past 15 years, increased seismic observations of the LLSVPs have motivated numerous geodynamical studies to understand their cause, many of which are briefly described in the preceding text. The first-order question pertains whether the LLSVPs are physical reservoirs of material that are compositionally distinct from the surrounding, background mantle. Although seismic observations increasingly hint at LLSVPs being compositionally distinct, it is unclear whether we can confidently exclude isochemical hypotheses for their cause, particular if such isochemical processes (e.g., plume clusters) may include passive compositional heterogeneity. For example, it is not inconceivable that LLSVPs may be caused by plume clusters that include a focusing of the oceanic crust and lithosphere that is being transported into them through the sides and dispersed through the top. If it is discovered that LLSVPs are indeed caused by distinct compositional reservoirs, then it is important to assess whether they are stable, long-lived structures that have resided at the base of the mantle for most of geologic time or whether they are active structures that rise and sink through the mantle. If they are stable, long-lived structures, it is then important to assess whether they are primordial structures with well-defined edges and tops or whether they are actively being created by the continual accumulation of oceanic crust. Furthermore, it is important to determine whether these structures are actively 'anchoring themselves' to the African and Pacific regions (as speculated due to hot spot and LIP correlations) or whether they are being guided by changing subduction patterns over geologic history.

7.11.5 Ultralow-Velocity Zones (ULVZs)

Ultralow-velocity zones are ~10 km thick regions at the CMB exhibiting seismic velocity decreases of order 10% or more, similar in scale and seismological contrast to the Earth's crust. Much like the crust is a key intermediary between the mantle and surface environment, it is reasonable to expect that ULVZs are also critically important as the frontier between the lowermost mantle and core. Since their discovery less than 2 decades ago (Garnero and Helmberger, 1996), studies of ULVZ have not advanced as far as other areas of deep-Earth science, and there are almost as many models as persons who have ventured to study these enigmatic features (e.g., Hernlund and Tackley, 2007). In the near future, it is hoped that integrated models of the thermal and chemical evolution of the CMB will lead to specific hypothesis tests that help to distinguish between potential origins for ULVZ. It is also important to incorporate ULVZ as a key outcome of processes operating at the CMB, and to explain their relationship to other major features in the deep interior.

7.11.5.1 Seismic Observations

The seismological detection and manifestation of ULVZ are well covered elsewhere (e.g., Lay, 2007); here, we give only a brief overview of the important features. *P*-waves are depressed by ~10% in ULVZ, and *S*-waves may slow by up to ~30%. The tops of ULVZ are relatively sharp, and modeling of the *ScP* phase interacting with ULVZ suggests a density increase of 5% (Reasoner and Revenaugh, 2000) and perhaps as much as 10% (Rost et al., 2005). Furthermore, ULVZ may exhibit internal gradients consistent with an increase in seismic-wave speed with depth (Rost et al., 2006).

ULVZs have been reported in many regions at the CMB, although it is not clear that they always exhibit the same characteristics (Thorne and Garnero, 2004). The Pacific is well oriented between seismic sources and stations for seismic phases turning in the CMB region and has received the most attention in seismological prospecting for ULVZ (Garnero and Helmberger, 1996). Recently, there have been reports of very large 'mega-ULVZ' in the Pacific region. In particular, the presence of an ~900 km wide ULVZ is reported beneath Hawaii (Cottaar and Romanowicz, 2012). An ~800 km wide ULVZ has also been detected inside a proposed gap (He and Wen, 2012) in the center of the mid-Pacific LLSVP (Thorne et al., 2013a). The Coral Sea region in the western Pacific hosts an ~700 km wide ULVZ, and the South China Sea region appears to contain several ~200 km wide ULVZs (Jensen et al., 2013). The northern and northeastern Pacific, on the other hand, does not appear to contain any detectable ULVZ (Rost et al., 2010). ULVZs have also been reported in a variety of other locations, such as Iceland (Helmberger et al., 1998), Africa (Ni and Helmberger, 2001), and the eastern Atlantic (Helmberger et al., 2000).

ULVZs have usually been described as isolated 'islands' of order several hundred kilometer width and 10–30 km thickness, rather than a global layer. Indeed, some portions of the CMB seem to lack any signature of complexity associated with ULVZ-like structures (e.g., Persh et al., 2001; Rost et al., 2010; Vidale and Benz, 1992). Thus, a ULVZ may truly be laterally discontinuous, rather than representing the thicker portions of an otherwise ubiquitous thin layer. However, earthquake sources large enough to be useful are not always impulsive enough to generate very high-frequency waves with ultrafine spatial resolution. Many seismic recordings lose resolving power at length scales under ~1 km at the CMB, such that

they cannot rule out the existence of a very thin (i.e., several 100 m) ULVZ-like layer spanning the entire CMB. Some seismic phases, such as *ScP*, can allow for finer-scale sampling and have been used to argue for the presence of <1 km layering of dense material on the underside of the CMB in some locations (e.g., core-rigidity zones or CRZ; Rost and Revenaugh, 2001), although it is not clear whether these are ubiquitous features of the CMB. Nuclear explosions are highly impulsive sources and also allow one to probe the CMB at fine scales. Ross et al. (2004) used Soviet nuclear test data to show that an ultrathin layer compatible with ULVZ-like properties exists beneath Siberia but is otherwise invisible to ordinary seismic waves. Although their coverage is limited, it is suggestive of an ultrathin layer that could be global in extent. The evidence therefore points to ULVZs as $\sim$10 km thick regions spanning several hundred kilometers in width, while other regions might host a very thin version of a ULVZ (<1 km).

7.11.5.2 Basic Dynamical Principles of a ULVZ Layer

If ULVZs primarily represent compositionally distinct material that is more dense than the overlying mantle, but less dense than the underlying core, then several conclusions may be drawn based on simple fluid mechanical principles. Firstly, in the absence of any external influences, a fluid layer of intermediate density sandwiched between layers of higher and lower density will tend toward a state that minimizes gravitational potential energy (GPE): a layer of uniform thickness. Therefore, lateral heterogeneity in the ULVZ structure at the CMB represents disturbances to this layer that can in principle provide information regarding the dynamical state of the CMB. The velocity of flow v_0 associated with viscous support of topography H_{ULVZ} on a thin compositionally distinct layer exhibiting a net density anomaly $\Delta\rho_{\mathrm{ULVZ}}$ is (e.g., Jellinek and Manga, 2004)

$$v_0 \approx \frac{\Delta\rho_{\mathrm{ULVZ}} g H_{\mathrm{ULVZ}}^2}{\eta_{\mathrm{ULVZ}}} \quad [1]$$

where g is gravitational acceleration and η_{ULVZ} is the dynamical viscosity of the ULVZ. Note that for a layer of distinct composition or phase constitution, η_{ULVZ} is probably different than the viscosity of the overlying mantle. Even if we allow for generous degrees of uncertainty in liquid iron viscosity at CMB conditions, the underlying core is essentially inviscid in comparison with the lower mantle; thus, the primary cause for lateral heterogeneity in ULVZ is expected to be deformation by viscous coupling to overlying mantle convection flows, which leads to thinning of the layer beneath downwelling mantle currents, and thickening of the layer beneath upwelling mantle currents (see **Figure 17**). Thus, the presence of ULVZ as an undulating dense layer at the CMB provides evidence for the presence of ongoing lower-mantle convection, and the distribution of thick and thin regions should reflect the pattern of mantle flow that prevails in the deep mantle (e.g., Sun et al., 2013).

Another important effect to consider is the relative topography generated on the upper and lower surfaces of a thin viscous intermediate layer such as the one described in **Figure 17**. Amplitudes on the upper surface will be greater if the density contrast is relatively small in comparison with the density contrast with the underlying layer. Because the shallow core has almost twice the density of the lower mantle, ULVZ-like structures of order 10% more dense than the overlying mantle should exhibit more dramatic topographic variations on the upper surface (by a factor of $\sim$8 for ULVZ density of 6 kg m^{-3}). Thus, topographic variations of order 10 km on the upper surface of ULVZ will be compensated by order 1 km topographic variations on the underside. In this way, ULVZ should protrude slightly into the core, by an amount that is already similar in magnitude to the longer wavelength amplitude of CMB topography produced by mantle convection processes (Olson et al., 1987). However, previous studies show that CMB topography produced by dense chemical layers is also sensitive to other factors, such as temperature and viscosity variations at a variety of length scales (Lassak et al., 2010).

7.11.5.3 Mechanisms and Hypotheses

Numerous models for ULVZ were proposed in the decade following their discovery and have already been reviewed (e.g., Hernlund and Tackley, 2007). Hypotheses linking ULVZ with mantle plume dynamics and geochemistry have also been reviewed (e.g., Jellinek and Manga, 2004). Here, we only consider the latest work on partial melting and composition anomalies, in addition to evolutionary scenarios that may lead to a combination of both mechanisms. As will be discussed in later sections, if a strong degree of secular cooling

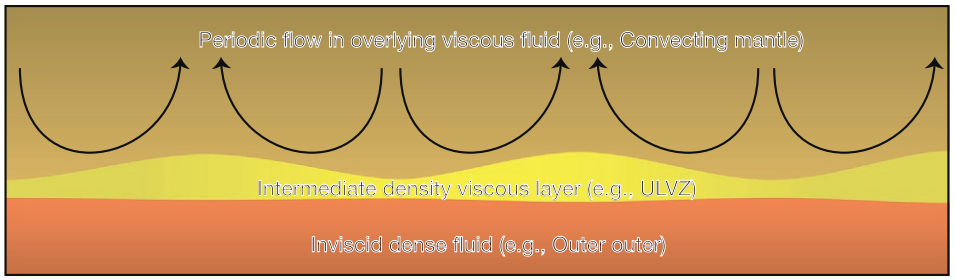

Periodic flow in overlying viscous fluid (e.g., Convecting mantle)
Intermediate density viscous layer (e.g., ULVZ)
Inviscid dense fluid (e.g., Outer outer)

Figure 17 Deformation of a thin layer of intermediate density (i.e., ULVZ) by flows in an overlying viscous medium (i.e., mantle). The underlying inviscid dense fluid (i.e., outer core) does not exert any shear stresses on the layer. In the absence of mantle flow, the topography would relax and flatten into a layer of uniform thickness. The amplitude of topography on the upper and lower surfaces of the layer partly reflects the density contrast between the layers; in this case, the density contrast with the overlying viscous fluid is assumed smaller than the density contrast with the underlying inviscid fluid.

is required at the CMB in order to sustain core convection and the geodynamo to satisfy paleomagnetic evidence, then it is probable that the base of the mantle was extensively molten in the past. The context of cooling of the base of the mantle from an extensively molten state to one that is near the solidus forms the basis of the 'BMO' hypothesis (Labrosse et al., 2007), which we also discuss.

7.11.5.3.1 Partial melt

Partial melting is an oft-cited hypothesis for explaining the properties of ULVZ, particularly the large amplitude and 3:1 ratio of shear-wave and compressional-wave velocity change reported in some seismological studies (Berryman, 2000; Hier-Majumder, 2008; Williams and Garnero, 1996). However, partial melting is a complex phenomenon and is not necessarily straightforward to apply as a mechanism for explaining ULVZ. Relevant questions regarding physical and chemical properties include the melt–solid density contrast (e.g., positive or negative), melt–solid partitioning of major and minor elements, changes in ULVZ viscosity, solidus and liquidus temperatures of material with ULVZ composition, and chemical reaction with the core, among others. All of these have an impact on the plausibility of the partial melt hypothesis. Following is a brief summary of ideas regarding partially molten ULVZ.

It has long been suspected that melts could become more dense than coexisting solid phases in equilibrium at lowermost mantle pressures (e.g., Agee and Walker, 1988; Akins et al., 2004; Hemley and Kubicki, 1991; Ohtani and Maeda, 2001; Rigden et al., 1984; Stolper et al., 1981; Suzuki and Ohtani, 2003; Suzuki et al., 1995). Melt–solid density inversions are well established in a variety of compositions at high P–T (Liu and Bassett, 1986) and in other geologic and planetary contexts (e.g., Agee and Walker, 1988; Warren, 1985). Melt–solid density is affected by two key parameters: (1) the partitioning of major elements with different atomic mass and (2) the volume change of melting. Volume changes upon melting exert an important control on the pressure dependence of melting. In the case of congruent melting, in which the solid produces a melt of identical composition, volume changes alone control the melt–solid density difference. In this case, melting obeys the Clausius–Clapeyron relation:

$$\frac{dT_m}{dP} = \frac{\Delta V}{\Delta S} \quad [2]$$

where T_m is the melting temperature, ΔV is the volume change, and ΔS is the entropy change (upon melting). The dT_m/dP slope is always proportional to the volume change, regardless of melt–solid composition differences, although the magnitude of the slope may differ from the one predicted by eqn [2] in the case of incongruent melting. If the melt has a larger volume, then melting will occur upon isothermal decompression, but if it has smaller volume, then melting is induced by isothermal compression. Although the volume change upon melting is usually positive in the shallow Earth for most mantle silicates, it decreases with pressure owing to the greater compressibility of the melt phase relative to solid phases, yielding a distinct convex shape in most melting curves. Even if the volume change of melting is positive, it is still possible for the melt to become more dense than the coexisting solid phases if enough heavy species preferentially partition into the melt relative to other species in solids. An important example is the relative partitioning of FeO and MgO, which is described by the distribution coefficient,

$$K_D^{Fe-Mg} = \frac{X_{FeO}^{sol} X_{MgO}^{liq}}{X_{FeO}^{liq} X_{MgO}^{sol}} \quad [3]$$

where X_i^j is the mole fraction of component i in phase j. If $K_D^{Fe-Mg} < 1$, then FeO preferentially partitions into the melt phase relative to the solid, and a mole of melt will be heavier than a mole of solid. This effect is likely to be important for melting of the lowermost mantle, for which experiments performed using a variety of compositions find that $K_D^{Fe-Mg} < 1$ (e.g., Andrault et al., 2011; Fiquet et al., 2010), and some have even suggested $K_D^{Fe-Mg} < 0.1$ (Nomura et al., 2011). Relative melt–solid density changes for the reported range of K_D^{Fe-Mg} thus fall in the range ~1–6%. Note that if FeO prefers to enter the melt phase, then the melting temperature of a rock should decrease with addition of FeO.

Partial melting of ordinary mantle material as it circulates through the lowermost mantle thermal boundary layer has been explored in 2-D mantle convection models (Hernlund and Tackley, 2007). When the solidus is less than the CMB temperature, the pattern of convection is modified relative to melt-free thermal convection depending upon the melt–solid density change and the degree of melt percolation from partially molten regions. The largest and thickest partially molten features are correlated with hot upwelling plumes. For a large retained melt fraction, a density decrease upon melting enhances the thermal buoyancy of plumes and tends to produce tall and narrow melt structures along upwelling plume axes. This kind of melt distribution is not consistent with observations of ULVZ. On the other hand, for a density increase upon melting, the partially molten region tends to flatten into thin broad features that are more compatible with seismic inferences of ULVZ. While they showed that ULVZ structures could be produced by dense melting, Hernlund and Tackley, 2007 also found that density anomalies required to produce ULVZ-like structures should drive percolation of melt on timescales that are small in comparison with the residence time of material circulating through the partially molten region. However, the simple percolation model was not able to capture other relevant dynamical effects, such as viscous stirring of ULVZ which may strongly influence melt migration (Hernlund and Jellinek, 2010).

7.11.5.3.2 Composition anomaly

As discussed previously, a partial melt explanation is often favored owing to the relative changes in shear-wave and compressional-wave speeds. However, ULVZs may exhibit a significant density difference relative to the surrounding mantle. If the density change is ~10%, then it is difficult to explain ULVZ as a result of partial melting alone. In the case where the density anomaly is ~10% or more, ULVZs are likely associated with (Fe-enriched) bulk compositions distinct from the mantle. Also, if ULVZs are distinct 'islands' of material without any thinner version of ULVZ between the thicker portions observed seismically, then ULVZs are more likely to represent a composition anomaly. Wicks et al. (2010) proposed that the density and seismic velocities of ULVZ could be explained by the

presence of iron-rich $(Fe_x,Mg_{1-x})O$ magnesiowüstite (with $x \sim 0.9$), a composition predicted by fractional crystallization of a BMO (Nomura et al., 2011) as will be discussed in the next section.

The case of ULVZ as a thin compositionally distinct layer was considered in a 2-D thermochemical mantle convection model (McNamara et al., 2010), which also included larger-scale LLSVP as distinct thermochemical piles. These models showed a tendency for ULVZ to accumulate at the edges of LLSVP (e.g., Hernlund and Tackley, 2007); however, they also showed that LLSVP breakup and merging events could be accompanied by transient behavior that could affect their distribution in any given seismological 'snapshot' of the mantle at present times. For example, merging of two piles would give rise to a transient episode in which the center of the new larger pile is transiently populated by ULVZ inherited from the former edges of the smaller piles, although these would eventually be pushed to the margins of the larger pile as the new flow regime becomes established.

7.11.5.3.3 Partial melting of compositionally distinct material

The simplest way to explain all the properties of ULVZ, as well as some potential variations in their seismic properties from one location to another, is that they represent compositionally distinct material that is at or near the melting temperature relevant for their particular composition. Thus, potential differences in *S*-wave speed anomalies between ULVZs in different locations could be explained by variations in the retained melt fraction inside ULVZs (e.g., Nomura et al., 2011). Since *S*-wave speeds are sensitive to melt fraction, even small differences in partial melt fraction could produce significant heterogeneity in ULVZ populations.

One remaining issue regarding the retention of partial melt inside at least some ULVZs is that even small melt–solid density differences will tend to drive percolation of the fluid through the pore spaces it occupies in the rock matrix. Simple models of ULVZ partial melting that included even a very slight degree of percolation exhibited rapid separation of melt from the solid phase (Hernlund and Tackley, 2007), which is an essential problem for partially molten ULVZ. A neutral melt–solid density difference may prevail; however, it would need to be nearly exact and persist over geologic timescales regardless of secular changes (e.g., CMB cooling, as will be discussed in later sections).

The retention of dense partial melt by viscous stirring of compositionally distinct ULVZ was considered by Hernlund and Jellinek (2010). The model is motivated by the fact that the topography of compositionally distinct ULVZ islands must be supported by viscous stresses arising from flow in the overlying mantle (as discussed earlier), with a stirring velocity v_0 given by eqn [1]. Viscous stirring sets up a dynamical pressure gradient inside the ULVZ that tends to drive melt toward the top of ULVZ, even when the melt is dense and would otherwise drain downward by gravity-driven percolation. The dynamical pressure gradient can become large enough to compete with the driving force for gravity-driven percolation of melt; thus, it must also be considered in the percolation problem. The ratio of the gravity force $\Delta\rho g$ to the dynamical pressure gradient is given by Hernlund and Jellinek (2010)

$$R = \frac{\Delta\rho g H_{\mathrm{ULVZ}}^2}{\eta_{\mathrm{s,\,ULVZ}^{v_0}}} \approx \left(\frac{\Delta\rho}{\Delta\rho_{\mathrm{ULVZ}}}\right)\left(\frac{\eta_{\mathrm{ULVZ}}}{\eta_{\mathrm{s,\,ULVZ}}}\right) \qquad [4]$$

where η_{ULVZ} is the net viscosity of partially molten ULVZ, $\eta_{s,\mathrm{ULVZ}}$ is the viscosity of melt-free ULVZ solids, and eqn [1] has been employed in the approximate expression for R. The value of $\eta_{s,\mathrm{ULVZ}}$ may be very different than the viscosity of the overlying mantle in this scenario because the relevant compositions should be different. Hernlund and Jellinek argued that if $\Delta\rho$ is smaller than $\Delta\rho_{\mathrm{ULVZ}}$ (i.e., ~1% as compared with ~10%) and the viscosity of partially molten ULVZ is less than melt-free ULVZ solids (i.e., $\eta_{\mathrm{ULVZ}} \ll \eta_{\mathrm{s}}$), then stirring-induced dynamical pressure gradients are probably at least 10–100 × stronger than the buoyancy force for gravity-driven percolation (i.e., $R < 0.1$). Their models suggest that for $R < 0.1$, a dense melt is readily driven upward in the ULVZ but that this mechanism also saturates owing to viscosity reduction with increasing melt fraction. A potential observable prediction of this model is that the melt fraction increases with height inside a viscously stirred ULVZ, such that seismic velocity should increase with depth as suggested by Rost et al. (2006). This phenomenon is similar to the focusing of melt migration toward a mid-ocean ridge axis owing to the deviatoric stresses induced by corner-flow (Spiegelman and McKenzie, 1987).

7.11.5.3.4 Basal magma ocean

The models discussed in the preceding text are relevant for the present dynamics and structure of ULVZ, but they do not directly address how ULVZ evolved in the context of the Earth's 4.5 Gy history. As will be discussed in a later section on core–mantle heat exchange, present evidence indicates that the CMB has cooled by more than ~1000 K since the Earth's formation, implying that the base of the early Earth's mantle was at temperatures well above the liquidus for any reasonable mantle composition (e.g., Gomi et al., 2013). Thus, the relevant long-term evolution problem at the CMB is not about melting ostensibly solid silicates to produce ULVZ, but rather involves gradual freezing of an extensively molten lowermost mantle from the early Earth to the present day. The proposal of such a BMO was primarily motivated by secular cooling of the CMB required to sustain a geodynamo since ~3.5 Ga (Labrosse et al., 2007).

Fractional freezing of an ancient BMO since 4.5 Ga provides a possible means for explaining many features of the lowermost mantle, including ULVZ and LLSVP, in addition to establishing a relationship between these features. In the simplest scenario, the BMO is born as a dense iron-rich gravitationally stable liquid layer underlying the solid mantle following the final stages of the Earth's accretion. Assuming that iron oxides depress the melting temperature of deep mantle rocks, fractional crystallization will evolve relatively iron-depleted rocks such that the residual BMO liquid is further enriched in iron. The net effect is that the density of the residual liquid increases and the melting temperature is further depressed. After some degree of Fe enrichment, the solids evolved from a BMO could become sufficiently dense relative to the average mantle that they could resist immediate entrainment into solid mantle flows, hence accumulating to form LLSVP. As fractional crystallization proceeds, the melt would also retain many volatile species whose enrichment would further depress the melting

temperature. With further cooling, the solid fraction of the residual BMO liquid increases to the point where it transitions to a mushy state, for which the solid grains form an interconnected skeleton and residual melt is trapped in the interstitial spaces. The resulting residual mushy mixture would be highly FeO-enriched, with a density anomaly of order ~10% relative to the overlying mantle, and could also remain partially molten as a consequence of fractionation. Thus, residual mush of a BMO may potentially explain all of the primary features of ULVZ.

There are still many questions to be resolved regarding the origin and evolution of a BMO, such as the mechanism of its formation in the early Earth. Several studies have considered the formation of a BMO following crystallization of a completely molten Earth (e.g., following a Moon-forming giant impact) from the mid-mantle, leading to contradictory results and interpretations (e.g., Andrault et al., 2012; Mosenfelder et al., 2009; Nomura et al., 2011; Stixrude et al., 2009; Thomas et al., 2012). While some of the differences between laser-heated DAC studies and shockwave experiments may be reconciled by adopting a revised model for shock heating (Wakabayashi et al., 2014), there is still disagreement on the precise value for equilibrium partitioning of iron between melt and solid (e.g., Andrault et al., 2012; Nomura et al., 2011).

7.11.5.4 Summary

The simplest present explanation for ULVZs is that they represent compositionally distinct material that is probably iron-enriched and this distinct material may or may not be partially molten. This carries important consequences for many processes operating at the CMB in the past and the present and will also play an important role in core–mantle chemical interactions, as will be discussed in a later section. A BMO is suggested by secular cooling constraints at the CMB, and if the required degree of cooling is as large as 1000 K (or greater), then there seems to be no way to avoid such a scenario. Our present challenge is to understand its formation and evolution, and to make predictions that can be further tested against observations. It is not clear that the BMO should have formed by crystallization of a completely molten Earth from mid-mantle depths, and alternative scenarios should also be considered.

7.11.6 Outermost Core Stratification

Much of the outer core exhibits properties that are broadly consistent with convection of a well-mixed isentropic liquid. Seismic models such as PREM (Dziewonski and Anderson, 1981) are able to fit a large volume of normal-mode and other data to an isentropic compression trend with little residual error. However, the data used in such models are not sensitive to departures from the average trend at the top and bottom of the outer core, and subsequent seismological work suggests some kind of anomalous layering at both boundaries. The nature of anomalous material in the outermost core is important because it influences exchange of both heat and matter across the CMB. Furthermore, such structural variations may reflect the consequence of mass transfer across the CMB over geologic timescales and are therefore an essential dynamical component of the CMB region.

7.11.6.1 Seismological Constraints on the Stratified Layer

Seismology is the most direct tool for documenting the presences or absence of outer-core stratification. The outermost core is most effectively sampled by *SmKS* waves (**Figure 2**) and mantle shear (*S*) waves with *SV* particle motion that convert to compressional (*P*) waves upon entering the core, bouncing $m-1$ times from the underside of the CMB and converting back to an *SV*-polarized *S*-wave upon exiting the core. All *SmKS* waves (SKS, *S2KS*, *S3KS*, *S4KS*, etc.) initiate at an epicentral distance corresponding with bottoming depths at the top of the outermost core. This is not the case for *PmKP* (including *PKP*) waves: the *P* velocity reduction going into the core results in *PmKP* waves that do not bottom shallower than the mid-outer core.

A number of seismic studies have used travel-time differences between successive *SmKS* waves to constrain the 1-D velocity gradient in the outer 50–400 km of the core (e.g., Eaton and Kendall, 2006; Garnero and Helmberger, 1995; Garnero et al., 1993a; Hales and Roberts, 1971; Lay and Young, 1990; Souriau and Poupinet, 1991; Tanaka, 2004, 2007; Tanaka and Hamaguchi, 1993). Past work is well summarized by Eaton and Kendall (2006). When compared to the PREM reference model (Dziewonski and Anderson, 1981), most studies provide evidence for velocity reductions between ~0.5% and 1% over the outermost 50–300 km. It is interesting to note that a velocity decrease is not consistent with a simple decrease in density at the top of the outer core without a larger than compensatory change in bulk modulus and appears to contradict Birch's law correlating density and seismic velocity (Birch, 1964). It is also noteworthy that a number of 1-D reference models suggest a distributed velocity reduction relative to PREM, for example, IASP91 of Kennett and Engdahl (1991), SP6 of Morelli and Dziewonski (1993), and AK135 of Kennett et al. (1995). The most recent seismological studies use *SmKS* at higher *m* than previous studies and suggest a velocity reduction distributed over 300 km thickness atop the outer core (e.g., Helffrich and Kaneshima, 2010).

Still, there are many potential errors and residual uncertainties in the seismic characterization of outermost core structure. Hales and Roberts (1971) measured their *SKKS*–*SKS* times without realizing that *SKKS* is 90° phase shifted from *SKS* (indeed, each underside CMB reflection produces an additional 90° phase shift). It was also found that the use of relative *SmKS* times is sensitive to heterogeneous structure at the bottom of the mantle, even if the ray paths are close together (Garnero and Helmberger, 1995; Garnero et al., 1993a; Tanaka, 2007). There is a historical pretext for skepticism about seismically inferred layering in the outer core since it was shown that the original F region of Bullen (1949) was in fact better explained by *D″* scattering (Doornbos and Husebye, 1972). And some of those who previously advocated for some kind of outermost core layering have now changed their opinion upon further examination (Alexandrakis and Eaton, 2010). Thus, it is useful to bear in mind that study of the outermost core layering is complex, and although many workers have

concluded that there is a low-velocity layer, others still disagree about the thickness and magnitude of the velocity contrast.

A sharp seismic reflection arising from liquid–liquid immiscibility such as that which occurs in the Fe–O–S ternary system at atmospheric pressures was sought using PmKP phases, but no compatible signature was found (Helffrich and Kaneshima, 2004). Thus, the variation in light elements in the outermost core probably does not cross a liquid–liquid miscibility gap. In principle, this kind of information can be used to constrain the potential compositions of an outermost core layer, provided that some ranges of composition produce immiscibility at the conditions of the CMB, while others do not. The existence of such immiscibility gaps at core pressures is still controversial, and further experimental work at CMB conditions is probably needed to better constrain thermodynamic models of metal–alloys atop the core.

7.11.6.2 Constraints from Geomagnetism

Geomagnetism also provides constraints on the density profile at the top of the core. In particular, secular variation of the geomagnetic field should be sensitive to the dynamical state of the outermost core. A number of research studies are motivated by the idea of inverting secular variations in the magnetic field for flow in the shallow core via the induction–diffusion equation (cf. Bloxham and Jackson, 1991):

$$\frac{\partial \vec{B}}{\partial t} = \vec{\nabla} \times \vec{v} \times \vec{B} + \frac{1}{\mu_0 \sigma} \nabla^2 \vec{B} \qquad [5]$$

where $\vec{B}$ is the magnetic field, $\vec{\nabla}$ is the gradient vector, $\vec{v}$ is the outer-core flow velocity, is its magnetic permeability, and σ is its electrical conductivity. A limitation of such models is that only the radial component of the magnetic field can be inferred, thus requiring additional simplifications to the flow regime in order to permit inversion for flow velocities. Incompressibility provides one reduction in the degrees of freedom, but one further assumption is also needed. Assumption of a purely toroidal flow, consistent with stratification, has long been thought to correspond to observed features of the secular variation such as stationary field saddle points (Whaler, 1980). Although it is possible to obtain such inversions for purely poloidal convective flow atop the core, opposite to the stratified layer assumption, the results are not easy to explain by physical mechanisms and suggest that the flow is probably toroidal in character (Beggan and Whaler, 2008).

The effects of stratification on dynamo generation and magnetic field characteristics have also been considered in the context of other planetary bodies. For example, models for Mercury's internal evolution yield thermal stratification of the outermost core; however, deep convection driven by the buoyancy evolved due to growth of a solid inner core can generate magnetic fields that are consistent with observations (Christensen, 2006; Wicht et al., 2007). Models for Saturn's interior that satisfy thermal evolution constraints imply the existence of a stably stratified layer that may contribute to the smaller dipole tilt of Saturn's field than is observed in other planets (Stevenson, 1980, 1982). Models that invoke a convecting region overlying a deeper stably stratified region have also been invoked to explain the unusual field morphologies of Uranus and Neptune (Stanley and Bloxham, 2004). Complete suppression of convection owing to stratification inside the cores of Mars, Venus, or other moons that no longer exhibit magnetic fields is also a possibility. Thus, stratification might be considered as a typical phenomenon within the broader context of general planetary evolution and reveals a large variety of potential behaviors. In such a context, the idea that the Earth's outer core may be partly stratified at the boundaries is not exotic or surprising.

Insights into the effects of the outermost core stratification might also be gained from numerical models of the geodynamo. Stanley and Mohammadia (2008) investigated fully 3-D numerical dynamos including a stably stratified outermost layer and homogeneous boundary conditions. They found a tendency for axial thermal winds to develop in cases including a stable layer, which drive a prograde (eastward) flow near the surface of the core in equatorial regions. This might present a problem because some features of the Earth's magnetic field exhibit a retrograde (westward) drift. Their result can be explained by the strong influence of rotation in the liquid outer core, which tends to align flow vorticity parallel to the axis of rotation. In the limit of vanishing viscosity, rotation causes the flow vorticity in the outer core to trend parallel to the rotation axis. While the Earth's core is not inviscid, viscosity is expected to be insignificant in comparison with rotation forces. In particular, the ratio of viscous to rotation forces is measured by the Ekman number ($E = \nu/2\Omega R_0^2$, where ν is the kinematic viscosity, Ω is the rotation rate, and R_0 is the core radius). For the Earth, E is estimated to be of order 10^{-15} or smaller. In this low-Ekman number regime, flows should exhibit a vorticity that is dominantly aligned with the axis of rotation, in which case the radial transfer of heat is greatly enhanced at the equator owing to the radial motions of the fluid at lower latitudes. At the poles, on the other hand, rotation suppresses radial motions and weakens radial advection of heat. This leads to significant differences in heat transfer through equatorial regions relative to the poles, which ultimately gives rise to a large-scale axial prograde thermal wind. The fact that such a wind is opposite in direction to Earth's westward drift, in spite of the evidence for the existence of a chemically stratified layer atop the Earth's core, presents a paradox and is the only finding that is contradictory to the existence of a stratified layer.

On the other hand, recent work on dynamo models with heterogeneous patterns of heat flux imposed at the top of the core suggests that the outermost core stratification may be necessary. These models use seismic tomography models at the bottom of the mantle in an attempt to impose geophysically consistent patterns of heat flux variations at the outer surface of the dynamo model. The resulting models of 'locked' dynamos, for which the velocity and magnetic field become stable, yield generated magnetic fields dominated by four concentrations of flux at high latitudes. These are similar to the present-day geomagnetic field, and their weak time variations also resemble features of the 400-year historical record (Gubbins et al., 2007). However, in order to obtain these kinds of locked solutions it is necessary to increase the influence of thermal conduction at the boundary. One way to produce this effect without artificially increasing the thermal conductivity is to allow for stratification to develop at the top

of the core by a combination of bottom heating and internal heat sinks, which suppresses the radial advection that would otherwise destabilize the high-latitude flux patches (Sreenivasan and Gubbins, 2008). Such dynamics suggest that the top of the Earth's core is only weakly convecting or is density stratified with no vertical motion.

The discrepancy between the aforementioned modeling interpretations might be reconciled in the future by using models that include both a stratified layer and heterogeneously imposed heat flux at the outer surface. It is likely that the influence of rotation on polar–equatorial differences in heat flux will compete with lateral heat flux variations imposed at the CMB by the pattern of deep mantle convection. Indeed, the existence of a westward drift in the Earth's field may offer a compelling constraint on the relative strength of heat flux differences arising from these two distinct mechanisms operating in the Earth's core.

Most dynamo models employ a codensity formulation for the composition and temperature fields (e.g., Braginsky and Roberts, 1995), but this simplifying assumption is not well suited for the physics of this situation, and double-diffusive models will likely be required to capture the relevant behaviors. The dynamics of core convection and the dynamo with a mixture of lateral heat flux variations and strong rotation have only begun to be explored in greater detail. Already dynamo models including a stratified layer atop the outer core are producing interesting results and suggesting new dynamical regimes (e.g., Nakagawa, 2011). Furthermore, generalization of the dynamo process to include fully double-diffusive dynamics involving both heat and composition driving forces while allowing for generalized behavior at the boundaries is still at a very young stage of development, and we look forward to seeing future developments in this field.

It has also been proposed that a stable layer could not be more than about ~100 km thick without being incompatible with the timescale for growth of features in the geomagnetic field that are thought to emerge via magnetic diffusion from the deeper core (Gubbins, 2007). This skin depth constraint may limit the thickness to even smaller thicknesses if the electrical conductivity is as high as indicated by recent mineral physics constraints (e.g., Pozzo et al., 2013). These models are only approximate, and improvements in integrating the spatial and temporal scale of flows atop the core are probably needed to establish a more firm constraint.

7.11.6.3 Stability of the Outermost Core Layers

Since core fluids are expected to exhibit a very small molecular viscosity, viscous entrainment is inefficient and the most effective potential mechanism for mixing a density stratified region at the boundary into the deeper core is by turbulent entrainment. The problem of turbulent entrainment of a layer exhibiting a gravitationally stable density gradient has received a lot of attention in the past owing to its importance in other dynamical settings, such as the shallow ocean. In the ocean problem, the surface is typically warmer than the deep waters, giving rise to a gravitationally stable thermal gradient called the 'thermocline.' Agitation by breaking waves and winds at the surface gives rise to turbulence that mixes the uppermost portions of this stable boundary layer, thus maintaining a dynamical balance that modulates the uppermost boundary layer of the world's oceans. The outermost core problem is more complex in several ways, particularly because Lorentz forces and double-diffusive dynamics also enter the picture. With these caveats in mind, the simpler turbulent entrainment problem may provide a rough guide for assessing the stability of outermost core layers; thus, we consider these arguments in lieu of treating the complete problem.

Entrainment and mixing of a gravitationally stable density gradient implies an increase in GPE that must be balanced. Conservation principles and experiments on turbulent entrainment of a stable layer indicate that the increase in GPE is compensated by a loss in kinetic energy density (KED) in the mixed layer (Turner, 1986), suggesting that the entrainment dynamics can be cast in terms of a nondimensional ratio between GPE and KED:

$$R_i = \frac{N^2}{(u')^2} = \frac{\partial \rho}{\partial z}\frac{g}{\rho (u')^2} \approx \frac{(d \log \rho) g L_M^2}{L_0 u_*^2}, \qquad [6]$$

where Ri is the 'Richardson number,' N is the buoyancy frequency, $\partial\rho/\partial z$ is the stabilizing density gradient, L_M is the thickness of a candidate mixed layer bounding the stratified region, L_0 is the total thickness of the stratified region, and u_* is the 'friction velocity' (typically assumed to be $\sim 10\times$ smaller than the ambient flow velocities in the convecting region).

Stability analysis for turbulent erosion of a density gradient shows that a portion of a density stratified layer adjacent to the convection region of thickness L_M will become mixed if $Ri < 1/4$ (Miles, 1961). For an ~1% density stratification distributed over 100 km atop the core, and a friction velocity of $10^{-5}\ \mathrm{m\,s^{-1}}$, one finds that only the lowermost L_m ~1 mm of the layer is subject to mixing by core flows according to this criteria. However, this length scale is so small that the relevant Reynolds number will also be quite small. For an ambient core flow of $u = 10^{-4}\ \mathrm{m\,s^{-1}}$, viscosity of $\mu = 10^{-3}$ Pa s, and density of $\rho = 10^4\ \mathrm{kg\,m^{-3}}$, the local Reynolds number ($Re = \rho u L_M/\mu$) is of order unity, which is 2–3 orders of magnitude smaller than values of Re that give rise to turbulent behavior. Unless core viscosity is as small as $\mu = 10^{-5}$ Pa s, then it appears that core flows may be incapable of generating the kind of turbulence at length scales small enough to allow for turbulent entrainment of a significant portion of the stable layer under the stated assumptions. These factors argue for long-term dynamical stability of such a layer against entrainment.

Another way of understanding the stability of the outermost core stratification is to consider that core flows are driven by lateral density perturbations of order 10^{-9} and may therefore be incapable of entraining and mixing material with a stabilizing density contrast which is seismically detectable and therefore significantly greater in magnitude (Stevenson, 1987). The limits of seismically detectable density perturbations are of order 10^{-3}; thus, any seismically detected stratified layer should exhibit six orders of magnitude stronger density contrast than the density perturbations that drive core flows. It is simply unreasonable to suppose that buoyancy forces driving convective flows could overcome opposing buoyancy forces more than a million times stronger in relative magnitude. Thus, any seismically visible stratified layer is likely stable against entrainment into the deeper core.

A more likely mechanism for downward encroachment of a stratified layer into the deeper core is via thermal diffusion of light alloys down the concentration gradient, which may advance the stratified frontier into the deeper core. For half-space diffusion, the frontier would advance downward in proportion to $\sqrt{4D\Delta t}$, where D is the atomic diffusivity and Δt is the elapsed time. For $D \approx 10^{-8}\ \mathrm{m^2\ s^{-1}}$ (Vočadlo et al., 2003), the layer would thus grow downward by ~100 km by half-space diffusion over the age of the Earth. This length scale is similar to what is observed by many seismological studies and suggests that the origin of the layer is caused by core–mantle reactions similar to those proposed by Buffett and Seagle (2010), although it may be too small to account for the much thicker layer inferred by Helffrich and Kaneshima (2010).

7.11.6.4 Composition of the Stratified Layer

One of the peculiar features of the stratified layer is a smaller seismic velocity than the average isentropic trend of the outer core. Clearly, the density must be smaller than the typical core material to provide for gravitational stability. Since the velocity varies as $\sqrt{K/\rho}$, where K is the bulk modulus, the stabilizing reduction in ρ must be more than compensated by a decrease in K in order to produce a negative velocity anomaly. A temperature anomaly is not expected to satisfy this condition, therefore the origin of the layer is likely to be explained by composition variations.

It is important to consider the kinds of general composition variations that could occur in such a layer. Consider the composition of the core fluid to be expressed as a vector as follows:

$$\vec{C} = [C_{\mathrm{Fe}}, C_{\mathrm{Ni}}, C_{\mathrm{O}}, C_{\mathrm{Si}}, C_{\mathrm{S}}, \ldots], \qquad [7]$$

where square brackets denote a vector string and C_{Fe}, C_{Ni}, C_{O}, C_{Si}, C_{s},... are the scalar concentrations of Fe, Ni, O, Si, S, ... The trace of $\vec{C}$ is unity, by definition. Now, we define the composition vector of the well-mixed isentropic convecting part of the core as $\vec{C}_0$ and decompose it into two parts:

$$\vec{C}_0 = \vec{C}_{M,0} + \vec{C}_{A,0}, \qquad [8]$$

where $\vec{C}_{M,0}$ represents the base metal solvent (Fe and Ni) and $\vec{C}_{A,0}$ represents the concentrations of alloying components. Similarly, we can express the concentration in the stratified layer as

$$\vec{C}_{\mathrm{S}} = \vec{C}_{\mathrm{M,S}} + \vec{C}_{\mathrm{A,S}}. \qquad [9]$$

With these definitions, we can be more explicit regarding the kinds of compositional variations that can occur in a stratified layer and how they are related to the origin of the layer itself.

One special case of an outermost core stratified layer is that of a proportional enrichment. A proportional enrichment occurs if the vector $\vec{C}_{\mathrm{A},0}$ is parallel to $\vec{C}_{\mathrm{A,S}}$ (i.e., $\vec{C}_{\mathrm{A},0} \| \vec{C}_{\mathrm{A,S}}$). Assuming that the alloys cause a density decrease in the metal, the gravitational stability condition implies that the magnitude of $\vec{C}_{\mathrm{A,S}}$ is greater than $\vec{C}_{\mathrm{A},0}$ Thus, we can write a proportional enrichment as $\vec{C}_{\mathrm{A,S}} = q\vec{C}_{\mathrm{A},0}$, where $q > 1$ is a scaling factor. A nonproportional enrichment occurs if $\vec{C}_{\mathrm{A},0}$ is *not* parallel to $\vec{C}_{\mathrm{A,S}}$, which is the more general case. Note that if the sound velocity of outermost core fluid is less than the adiabatic trend, as indicated by many seismic studies, then we require that

$$\frac{K_{\mathrm{S}}\left(\vec{C}_{\mathrm{S}}\right)}{K_{\mathrm{S}}\left(\vec{C}_0\right)} < \frac{\rho\left(\vec{C}_{\mathrm{S}}\right)}{\rho\left(\vec{C}_0\right)} < 1, \qquad [10]$$

where K_{S} is the adiabatic bulk modulus, ρ is the density, and the last inequality results from the requirement of gravitational stability.

The condition expressed by [10] may be easier to satisfy in the case of a nonproportional enrichment, since it allows for more degrees of freedom. One of the main issues with proportional enrichment is that a velocity decrease is not usually considered plausible, unless one invokes strongly nonideal mixing (Helffrich, 2012). On the other hand, recent work on mixing of Fe–FeO in the Earth's core suggests that addition of oxygen alone may lower seismic velocity (Komabayashi, 2014). In this case, a proportional enrichment of oxygen could explain a velocity decrease, but such a profile predicts seismic velocities deeper in the outer core that are much slower than observed. Thus, a proportional enrichment alone may be unlikely to explain the usual observations.

7.11.6.4.1 Alloy incompatibility from inner-core growth

If the origin of the outermost core stratified layer is due to incompatible alloys that were expelled by growth of the inner core (e.g., Fearn and Loper, 1981; Helffrich and Kaneshima, 2013), then in the simplest circumstances, we would expect a proportional enrichment in those alloys. A similar situation holds for a stratified layer formed by barodiffusion of light alloys from greater depth and accumulating at the surface of the core (Braginsky, 2007; Gubbins and Davies, 2013). However, another possibility is that $\vec{C}_{\mathrm{A},0}$ contains more than one significant component, one of which is more incompatible than others in the inner core. For example, Alfé et al. (2002) suggested that oxygen is less compatible than sulfur and silicon at inner-core conditions. In such a case, light alloy accumulation may not necessarily give rise to proportional enrichment. Such a mechanism for producing a nonproportional enrichment can in principle be predicted and tested.

A separate problem with light element accumulation due to inner-core crystallization is that the inner-core boundary layer that expels buoyant-incompatible light alloys is expected to be of order 1–10 cm thick (e.g., Deguen et al., 2007), and upwelling currents transporting light alloy-enriched material upward will exhibit a similar length scale. For diffusivities of order $10^{-8}\ \mathrm{m^2\ s^{-1}}$, such an upwelling tendril would have a characteristic diffusion time ranging from several hours to a week, thus becoming quickly diluted. For typical outer core flow velocities ($\sim 1 \times 10^{-4}\ \mathrm{m\,s^{-1}}$) a transit time of order ~1000 years is required to rise from the inner core to the CMB, more than $\sim 10^4$ diffusion times for an upwelling of length scale 10 cm. This estimate is a lower bound because it assumes that such an upwelling current rises directly upward, which is not likely in a convecting fluid subject to the relatively strong influence of rotation. Furthermore, the stretching and mixing by the turbulent environment of outer core convection would

further shorten the timescales over which such structures could remain coherent. Thus, it appears difficult to accumulate buoyant compositionally distinct fluid atop the core.

Another problem is that, while buoyancy sources such as light alloy expulsion from inner-core growth could very well be an important part of driving core convection (Braginsky, 1963), it has already been mentioned that the density fluctuations associated with whatever buoyancy sources are present in the Earth's outer core exhibit lateral density fluctuations of order 10^{-9}, at least six orders of magnitude too small to produce anything that could be seismically detected (Stevenson, 1987). Thus, it is difficult to imagine any driving force for core convection to account for seismological observations of any kind.

7.11.6.4.2 Core–mantle reaction-induced stratification

We discuss core–mantle reactions in greater detail in a later section; thus, we only briefly mention the relevant background for understanding outer-core stratification arising from core–mantle reactions. Many recent experimental and theoretical studies have found that typical mantle compositions will be out of equilibrium with outer-core metal containing a maximum of ~10 wt.% light alloys. The expected direction of the reaction, according to recent studies (Asahara et al., 2007; Frost et al., 2010; Ozawa et al., 2009), is for oxygen to dissolve into the metal. Asahara et al. (2007) initially used this result to argue for iron depletion of the bottom of the mantle, because they did not consider the possibility that the metal reaction products would be gravitationally stable atop the core, thus limiting the extent of the reaction. Indeed, according to seismic observations, the base of the mantle appears Fe-enriched, not Fe-depleted (see ULVZ discussion). Fe enrichment at the base of the mantle should increase the extent of disequilibrium between ambient core fluid and the base of the mantle, requiring even more oxygen atop the core to bring the system closer to equilibrium.

One of the main advantages of core–mantle reactions for explaining seismic observations is that it allows more degrees of freedom to explain the low-velocity anomaly because $\vec{C}_{A,0}$ is *not* parallel to $\vec{C}_{A,S}$. Buffett and Seagle (2010) presented a simple model for a stratified layer produced by core–mantle reactions in the Fe–O system. Their model did not include the diffusion couple on the mantle side, nor did it include a large degree of expected secular cooling of the CMB, and the stratification was consequently weak and subject to significant degrees of entrainment into core flows. Nevertheless, the model exhibits many first-order characteristics that are unlikely to change with more detailed models. As mentioned in the preceding text, diffusivities of up to 10^{-8} m^2 s^{-1} lead to downward encroachment of the layer on the order of ~100 km, a likely outcome of any model employing similar diffusivities. However, one of the main difficulties with this model is that the diffusivities may be too small to account for the much greater thicknesses inferred by Helffrich and Kaneshima (2013), and this difference remains to be explained. Nevertheless, recent thermodynamic models of mixing in the Fe–FeO system (Komabayashi, 2014) provide further support for the Buffett and Seagle mechanism and may provide the simplest explanation for seismological inferences.

7.11.6.5 Summary and Perspective

Both the existence and origin of a stratified layer atop the core are ongoing topics of investigation, and neither should be considered adequately resolved. Highly skilled seismologists have come to completely different conclusions regarding the robustness of inferred layer structures, while mineral physics constraints still exhibit considerable uncertainties. Geomagnetic arguments are confounded by nonuniqueness. The outermost core stratification remains one of the most important first-order problems in deep-Earth geophysics, yet it appears that more work is needed before we can move toward a consensus.

Important clues may still be found and leveraged for the purposes of interpretation. For example, the F layer at the bottom of the outer core also exhibits a negative velocity contrast relative to the adiabatic trend and is probably better resolved than the outermost core stratification (Souriau and Poupinet, 1991; Zou et al., 2008). However, for stability, the F region must exhibit a higher density than the isentropic part of the convecting outer core. If such a layer can be explained by proportional depletion in light alloys (i.e., $\vec{C}_{A,S} = q\vec{C}_{A,0}$ and $q < 1$), then any outermost core stratification that produces a negative density and velocity anomaly likely has an origin distinct from proportional enrichment (Hirose et al., 2013).

7.11.7 Core–Mantle Heat Exchange

CMB heat flow provides a vital link for understanding the fundamental thermal, chemical, and magnetic evolution of Earth's interior. Recent years have witnessed a steady increase in estimates for CMB heat flow, due in part to increased estimates of CMB heat flow based upon hot spot swell buoyancy flux constraints (Labrosse, 2002; Mittelstaedt and Tackley, 2005; Zhong, 2006), in addition to proposed constraints arising from the PPv double-crossing (Hernlund et al., 2005; Wu et al., 2011) and upward revisions in estimates of core thermal conductivity from a combination of several ab initio modeling approaches, laboratory experiments, and recognition of saturation resistivity in the geophysical context (de Koker et al., 2012; Gomi et al., 2013; Pozzo et al., 2012) that place strong constraints on CMB heat flow as far back as ~3.5 Ga. Until these recent developments in understanding CMB heat flow, it was common to think of CMB heat flows of ~5 TW (e.g., Buffett, 2002); however, at the present time, most lines of evidence argue for a CMB heat flow more than 10 TW. This major upward revision in CMB heat flow carries important consequences for the evolution of the Earth's deep interior, requiring ~1000 K of secular cooling at the CMB and extensive melting of the lowermost mantle in the early Earth. In the following sections, we briefly review each of these recent contributions to understanding CMB heat flow.

7.11.7.1 Plume Heat Flow

Volcanic hot spots are associated with topographic swells, or upward deflections of the Earth's surface in the region surrounding the volcanic center. According to models of hot spots involving a deep-seated mantle plume origin, these

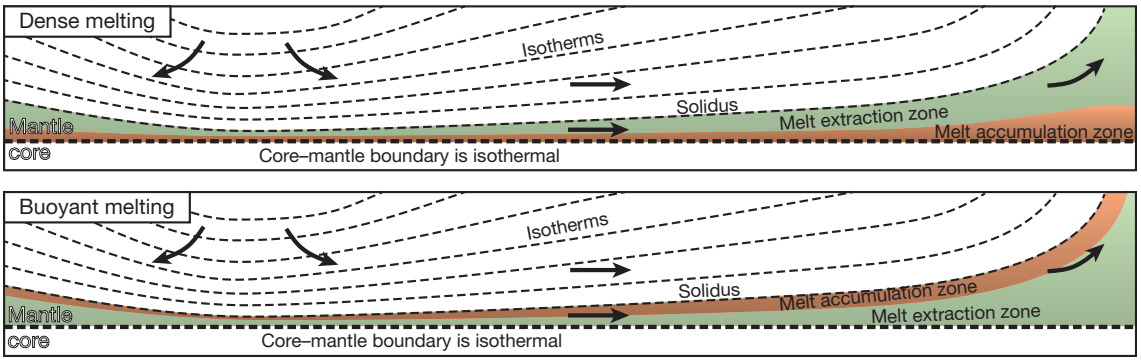


Figure 18 Cartoon of melting scenarios in the lowermost thermal boundary layer, assuming that the CMB is hotter than the mantle solidus. The top and bottom panels show the outcome of scenarios in which the melt is more or less dense than solids, respectively. Arrows represent the mantle flow with downwelling on the left side and upwelling (e.g., a plume) on the right. Melt percolates up or down depending on the melt–solid density difference, leaving melt-depleted rock (green) above or below a melt-enriched layer (orange).

swells are dynamical topography generated in proportion to the upward buoyancy flux of hot material as it rises through the *upper* mantle beneath the lithosphere. Employing this mechanism yields estimates ranging from 41 to 55 Mg s^{-1} buoyancy flux globally (Davies, 1988; Sleep, 1990). If one assigns temperatures and thermal expansivity values to these plumes, such buoyancy fluxes correspond to roughly 2–4 TW advected heat flow carried upward in the *upper* mantle. This order of magnitude of heat flow associated with swells has sometimes been used to argue for a deep, rather than shallow, origin for hot spots (**Figure 19**).

It has sometimes been assumed that the heat flux carried by plumes in the upper mantle is a reliable estimate of the CMB heat flow, giving rise to broad uncertainties. However, mantle convection studies argue that upper-mantle plume heat flux underestimates the heat flow by a factor of 3–5. While all recent studies agree that upper-mantle plume heat flow is much less than the CMB heat flow, the precise reasons are still debated. **Figure 19** gives a brief summary of some arguments. Heat is carried through the mantle by upwellings and downwellings. One view holds that heat flow is dominated by sinking lithospheric slabs in the shallow mantle and by upwelling plumes in the deeper mantle, with the heat carried by hot spot-compatible plumes in a compressible mantle decreasing dramatically upon ascent owing to differential adiabatic cooling, conductive heat loss, and other factors (Zhong, 2006). Another view holds that some plumes may be very weak (Labrosse, 2002), assume the form of transient thermals (Jellinek et al., 2003), or represent passively entrained portions of the lower boundary layer (Romanowicz and Gung, 2002), and in such cases, they may not produce a significant or easily identifiable topographic swell even if they reach the upper mantle.

While a significant underestimation of CMB heat flow by hot spot swells is indicated by many studies, the precise conversion factor between these quantities is not known to a high degree of accuracy. Using plume buoyancy flux in addition to other constraints such as upper-mantle temperatures, calibrated by results from 3-D spherical shell thermochemical convection models, Zhong (2006) proposed CMB heat flow in the range 12.6–14.2 TW. While other studies employ different kinds of models, their results are also consistent with such a large conversion factor between the shallow plume heat flow and CMB heat flow (e.g., Mittelstaedt and Tackley, 2005). Thus, the CMB heat flow in excess of 10 TW is compatible with hot spot swell constraints, and the most detailed models suggest values close to 13 TW.

7.11.7.2 CMB Heat Flow Estimates from *D″* Conductivity and Thermal Gradients

Because radial heat advection vanishes at the CMB, conduction is the only means of passing heat from the top of the core into the very bottom of the mantle. Therefore, one way of estimating CMB heat flux is by direct application of Fourier's law: multiplying the lowermost mantle thermal conductivity by an estimate for the average thermal gradient and using a factor of 1.52×10^{14} m^2 for the CMB area to convert from heat flux to total heat flow. This is in principle a very straightforward procedure, except that both the thermal conductivity and average *D″* geothermal gradient are highly uncertain. Direct determinations of absolute temperature in the deep mantle from seismic velocity variations (e.g., Kawai and Tsuchiya, 2009) are probably ill-posed owing to the accumulation of systematic errors in seismic velocity profile, poorly constrained dependence of seismic velocity upon temperature (and unknown composition), and the influence of independent factors such as variations in elastic anisotropy. The recent discovery of a Pv–PPv phase change has possibly improved our state of knowledge, and when combined with more complete boundary layer and convection models, it may be possible to narrow the range of uncertainty considerably, although the results are model-dependent. While new experimental and theoretical/computational constraints on *D″* thermal conductivity have also emerged, there is still residual uncertainty and thus these estimates should be considered a work in progress. Nevertheless, as we will explain in the succeeding text, the emerging range of CMB heat flow estimates based on recent developments are beginning to converge with other estimates such as plume heat flux (as discussed in the previous section) and core heat flow (discussed in the next section). The mutual consistency of these constraints is a potentially promising development for understanding the dynamics and evolution of the deep Earth.

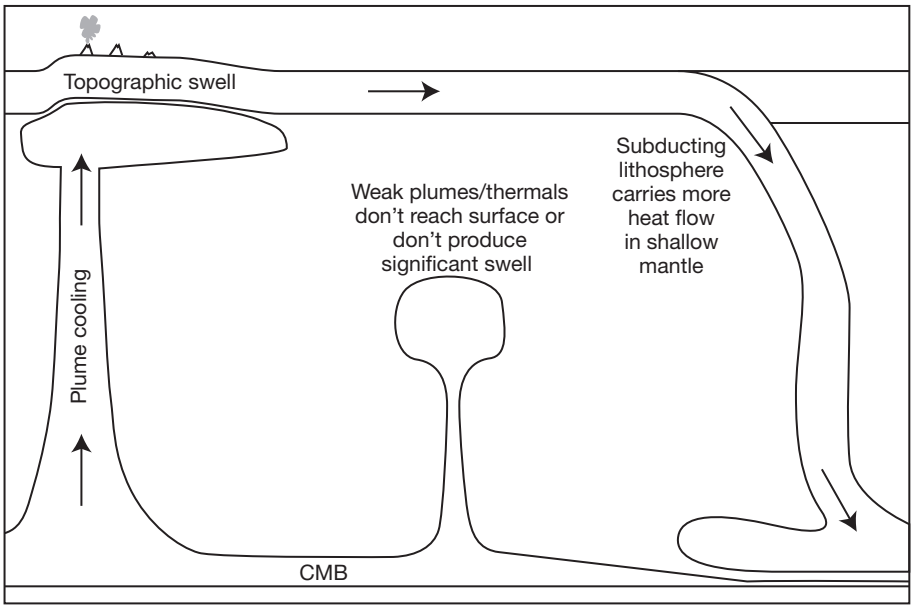

Topographic swell
Subducting lithosphere carries more heat flow in shallow mantle
Weak plumes/thermals don't reach surface or don't produce significant swell
Plume cooling
CMB

Figure 19 Illustration of some commonly cited reasons that plume heat flux in the upper mantle significantly underestimates CMB heat flow. Plumes may cool more rapidly upon decompression at high temperature than surrounding material in a compressible mantle and lose heat by conduction. Some have proposed weaker plumes or thermals that either do not reach the shallow mantle or do not produce a measurable topographic swell. Some views hold that downwelling slabs carry much of the advected heat flow in the shallow mantle, while plumes carry more heat flow in the deeper mantle.

7.11.7.2.1 Estimates of mantle thermal conductivity

The thermal conductivity of D'' is an important parameter, and is critical for obtaining heat flux estimates from the geothermal gradient. For many years a central value of 10 W m^{-1}K^{-1} was assumed by many workers (e.g., Stacey, 1992), although enough was known to establish at least a factor of 2 uncertainty. By modeling laser heating power versus temperature in the DAC for materials with different conductivities, Manga and Jeanloz (1997) estimated a deep mantle conductivity in the range 5–12 W m^{-1}K^{-1}, with trade-offs in estimates of CMB temperature, similar to models proposed by Goncharov et al. (2007). Recent estimates for Pv-bearing rock fall in this range, including recent experimental measurements and *ab initio* calculations that yield conductivities in the range 9–11 W m^{-1}K^{-1} at the CMB (Manthilake et al., 2011; Ohta et al., 2012). The narrowing in the range of estimates is hopeful; however, these results await further confirmation and a number of other experimental and modeling efforts are currently under way that may yield significantly different values.

The aforementioned estimates are for the lattice thermal conductivity; however, it has also been proposed that radiative effects could be important in the Earth's deep mantle, resulting in a higher effective conductivity (Clark, 1957; Hofmeister, 1999). This radiative conductivity contribution is sensitive to absorption in the near-infrared part of the spectrum, where much of the photon transmitted power lies for mantle temperatures. Early experiments showed that near-infrared absorption increased with pressure, mitigating the effects of increased radiative power at elevated temperatures (Mao and Bell, 1972). Recent measurements of absorption for lower-mantle phases up to CMB pressure are also consistent with a limited radiative effect on lower-mantle conductivity, of order 10% or less compared with the lattice conductivity (Goncharov et al., 2008). Thus, the radiative contribution is usually ignored, since there is little evidence to support a significant effect at depth. However, this still represents another source of uncertainty that must be considered in boundary layer models of the D'' region, and estimates based on lattice conductivity alone must be considered a lower bound.

7.11.7.2.2 Estimates of temperature change and boundary layer thickness

One possible method of obtaining an estimate of the adiabatic temperature drop across the D'' thermal boundary layer ΔT_S is to subtract the extrapolated potential temperatures of the core and mantle at the CMB, anchored by transition zone temperatures from the olivine–spinel transition and freezing temperature estimates for impure iron at the inner-core boundary (Williams, 1998). However, such a procedure yields factor of 2 uncertainties and does not sufficiently constrain the actual quantity of interest: the gradient. An average thermal boundary layer thickness L can additionally be assumed, such that the average gradient is assumed to be $\Delta T_S/L$. However, there are no reliable ways of estimating L, particularly in the presence of complexities such as large-scale chemical heterogeneity, and uncertainties in this parameter also exceed a factor of 2. L may also vary by more than two orders of magnitude in D'' from one region to another, particularly under the influence of thermochemical convection. Thus, the product $\Delta T_S/L$ yields combined uncertainties of a factor of 4 or greater. If one adds an additional factor of 2 uncertainty in thermal conductivity, the uncertainty in heat flow determined by this approach is more than an order of magnitude, which is clearly unacceptable.

7.11.7.2.3 Geothermal gradients and postperovskite

Some phase changes offer reliable determinations of temperature if their depth (pressure) can be tightly constrained by seismological observations of discontinuities caused by changes in elastic properties and density across the transition. Thus, the Pv–PPv transition located in the deep mantle may be seen as quite fortuitous for revealing temperatures in the deep mantle. However, the position of the phase boundary and the Clapeyron slope are still not confidently determined owing to uncertainties in pressure standards used in high-pressure experiments. While ab initio models have improved dramatically in recent years, the interaction potentials used in the calculations are necessarily based upon approximations to full solutions to Schrodinger's equation, and it is not easy to obtain robust error estimates. Therefore, the Pv–PPv phase transition alone exhibits many uncertainties, unless one regards a particular experimental pressure standard as being robust and geophysically consistent, or a specific ab initio technique as being superior to all other constraints.

The Pv–PPv transition may help to constrain ΔT_S, but the procedure is not straightforward. Some attempts have been made to construct a whole mantle geotherm based in part upon the temperature of the Pv–PPv transition at a depth of 2700 km and have been used to estimate the magnitude of the mantle adiabatic gradient (e.g., Ono, 2008). However, as discussed previously, the most straightforward and consistent interpretation of a Pv–PPv origin for the D'' discontinuity is in the circum-Pacific, inside deeply subducted lithosphere that is ~700 K cooler than surrounding mantle (e.g., Hernlund and Labrosse, 2007). Therefore, the temperature at 2700 km given by the Pv–PPv transition more likely refers to the temperature of cold subducted slabs, which is not representative of ambient adiabatic mantle, and therefore significantly underestimates the adiabatic gradient in the mantle.

7.11.7.2.4 Geothermal gradient constrained by a double-crossing

A Pv–PPv double-crossing is only possible if temperature gradients at the CMB exceed the gradient along the Pv–PPv phase boundary, which is given by $\rho g/\Gamma$ (where ρ is the density, g is the acceleration of gravity, and Γ is the Clapeyron slope). The product ρg in D'' is well constrained by seismological models ($\rho g \approx 5.5 \times 10^4$ Pa m^{-1}), while Γ must be greater than about 8 MPa K^{-1} in order for a Pv–PPv double-crossing to be geophysically compatible with other constraints (Hernlund and Labrosse, 2007). Using the range 9–13 MPa K^{-1}, we obtain Pv–PPv phase boundary gradients in the range 4.2–6.1 K km^{-1}.

This constraint is only strictly a lower bound on the geothermal gradient and is only applicable if and where a Pv–PPv double-crossing occurs but may be highly useful because it offers the first direct inference of geothermal gradients just above the CMB in a way that can be constrained directly by laboratory experiments. Furthermore, this constraint is most useful when applied to cold ponded downwellings residing just above the CMB, since these are the locations where most of the heat will be absorbed from the core into the mantle by heat conduction. Denoting the fraction of area of the CMB overlain by a Pv–PPv double-crossing as f_{dc}, a lower bound on total CMB heat flow Q_{cmb} can be written as

$$Q_{cmb} > f_{dc} 4\pi r_{cmb}^2 k \frac{\rho g}{\Gamma}. \quad [11]$$

The challenge in applying this kind of estimate is seismic coverage, which limits our ability to reliably establish an appropriate value for f_{dc}. For reference, if $f_{dc} = 1$ and $k \approx 10$ W m^{-1} K^{-1} (e.g., Manthilake et al., 2011; Ohta et al., 2012; Stacey, 1992), then the aforementioned range of estimates would yield Q_{cmb} >6.4–9.3 TW. Of course, on the basis of seismic observations, we expect that f_{dc} is less than unity, which would reduce this lower bound estimate, but just how much less is not well constrained given insufficient seismic coverage of D''. The actual geothermal gradient could also be much steeper than the phase boundary (see discussion in the succeeding text), corresponding to a larger Q_{cmb}. Therefore, uncertainties exist that would either decrease or increase our estimate of Q_{cmb} based on a Pv–PPv double-crossing, illustrating the trade-offs in utilizing this kind of constraint.

Another complexity in applying the Pv–PPv double-crossing as a constraint on geothermal gradients at the base of the mantle is that material downwelling through a Pv–PPv transition will absorb latent heat, which causes a further steepening of the geothermal gradient above $\rho g/T$ (see **Figure 20**). Buffett (2007) first estimated this effect for a univariant transition and revised the double-crossing constraint as follows:

$$Q_{cmb} > f_{dc} 4\pi r_{cmb}^2 k \left(\frac{\rho g}{\Gamma} + \frac{v_z L}{\kappa C_p} \right), \quad [12]$$

where v_z is the downwelling velocity through the lower transition, L is the latent heat, κ is the thermal diffusivity, and C_p is the specific heat. $L = T\Delta S$, where ΔS is the entropy change for the transition, and can be estimated from the P–T slope of the Pv–PPv phase boundary and volume change via the Clausius–Clapeyron relation. However, it is difficult to estimate v_z just above the CMB based upon dynamical considerations, as there are many uncertainties in viscosity and other parameters (Buffett, 2007). Another approach is to use solutions to the advection–diffusion equation in the boundary layer to estimate v_z/κ. Assuming a first-order variation in downwelling velocity proportional to height z above the CMB such that $v_z = \dot{\varepsilon} z$ (where $\dot{\varepsilon}$ is the strain-rate), it can be shown (e.g.,

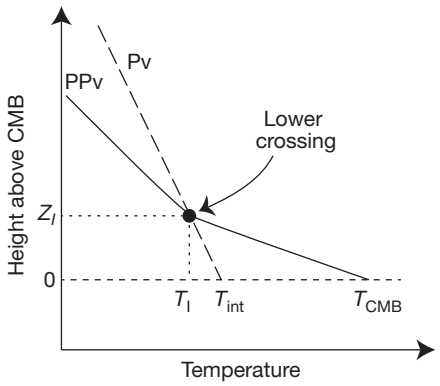


Figure 20 Steepening of the geotherm (solid line) below a lower-crossing of the Pv–PPv phase boundary (dashed line) due to absorption of latent heat.

Hernlund, 2010) that a boundary layer approaches steady conditions such that $\dot{\varepsilon}/\kappa = 2/z_0^2$ over timescales of $1/\dot{\varepsilon}$, where z_0 is a measure of the boundary layer thickness. For the range $z_0 \approx 100$–150 km, compatible with the Pv–PPv double-crossing in a downwelling scenario, one obtains $\dot{\varepsilon}/\kappa \approx 1 \times 10^{-10}\mathrm{m}^{-2}$. Thus, v_z/κ at the height of the lower discontinuity z_l above the CMB is $\dot{\varepsilon} z_l/\kappa \approx z_l \times 10^{-10}\mathrm{m}^{-2}$. For $C_p = 1300$ J K^{-1} kg^{-1} and $L = 100$ kJ kg^{-1}, the term $v_z L/\kappa C_p$ in eqn [12] is $\sim$1 K km^{-1} for $z_l = 10$ km and increases linearly with z_l (although the linear approximation becomes less reliable further above the CMB). For comparison, latent heat release for a PPv → Pv lower-crossing at heights of ~30–50 km above the CMB could steepen the thermal gradient by an amount similar in magnitude to the Pv–PPv phase boundary gradient itself, doubling the implied heat flux bound.

The addition of a two-phase region increases the implied geothermal gradient even further than steepening due to latent heat absorption (Buffett, 2007). However, if the thickness of the Pv–PPv coexistence region is large, then the effects of latent heat become less important in numerical solutions, perhaps due to the distribution of latent heat effects over a larger depth range in addition to the combined effect of latent heat release at the Pv → PPv upper-crossing which tends to exert an opposing influence on the shape of the geotherm (Hernlund, 2010). In any case, when the phase transition thickness is very wide, it is difficult to apply the simpler formulas like eqn [12], and more complete numerical solutions are required to assess the contributions of latent heat.

One method to obtain a direct estimate of the thermal gradient from the double-crossing without requiring any constraints upon the latent heat effect is to assume the position and slope of the Pv–PPv phase boundary, as well as the temperature of the CMB. With this information and an observed depth for the lower-crossing z_l, one can directly calculate the geothermal gradient below the double-crossing by computing $(T_{\mathrm{cmb}} - T_l)/z_l$, where T_l is the temperature at the lower-crossing (see **Figure 20**). This is clearly the most direct approach (van der Hilst et al., 2007); however, it requires accurate knowledge of the phase boundary and CMB temperature, both of which are uncertain by as much as several hundred K; therefore, this constraint is not very reliable at the present time. It is also possible to fit models for the entire geotherm to both the upper-crossing and lower-crossing to obtain an estimate of heat flux using either analytic (Lay et al., 2006) or numerical approaches (Hernlund, 2010). While these solutions exhibit trade-offs with the shape of the geotherm, it is difficult to utilize the additional information as another constraint, and in practice one still encounters similar issues with respect to uncertainties in the phase boundary and CMB temperature.

7.11.7.2.5 CMB heat flux from seismic tomography and boundary layer models

In order to utilize a PPv double-crossing as a constraint upon total CMB heat flow, it is necessary to extrapolate the geothermal gradient from regions where a double-crossing occurs to the remainder of the CMB region. These methods will hopefully yield more relevant constraints on CMB heat flow from considerations of conductivity and thermal gradient variations in the lowermost mantle, and their development is also vital for providing a bridge to other areas of lower-mantle research.

Seismic tomography potentially offers a wealth of information about variations in temperature above the CMB, which could in principle be used to constrain the CMB distribution of heat flux. One possibility is to assume a simple proportionality between seismic-wave speeds in the deepest mantle and CMB heat flux, and this kind of approach is often employed as boundary conditions for dynamo models that seek to relate mantle structure to features of the geomagnetic field (e.g., Aubert et al., 2008; Glatzmaier et al., 1999; Willis et al., 2007). However, this requires assigning a physical value for heat flux to seismic velocity, for example, by defining the upper and lower limits of heat flux using constraints from a Pv–PPv double-crossing, numerical convection models, boundary layer models, or some combination of these.

Here we give a simple illustration of using a tomographic model to extrapolate heat flux to the entire CMB, by assuming a linear dependence between heat flux and seismic velocity. It is important to realize that the minimum heat flux will always be negligible relative to any reasonable maximum heat flux. This is because the lowest heat flux corresponds to the highest temperatures, which will be nearly the same as the CMB temperature (Nakagawa and Tackley, 2008). Therefore, one may build a rough CMB heat flux map by linearly transforming a seismic tomographic model such that the lowest seismic velocities are nil and the highest seismic velocities are unity. The map may then be subsequently scaled by a suitable value for the maximum heat flux, yielding an estimate for CMB heat flux in all locations. The PPv double-crossing is a useful constraint, because it should occur beneath cold downwellings and therefore gives an estimate for the maximum heat flux.

The choice of tomography models is also important. While *S*-wave models have often been used for this purpose, they should be highly sensitive to variations in composition or PPv fraction and are therefore not a good proxy for temperature. In fact, this known sensitivity underlies much of the recent work on D'' discontinuities and LLSVP, as discussed previously. *P*-wave models, on the other hand, should be more sensitive to temperature than to composition or PPv variations and are thus preferred for this purpose (e.g., Trampert et al., 2004). While *P*-wave models are thought to be less reliable owing to relatively poor coverage in the deep mantle (compared with *S*-wave models), the growth in over all data coverage with time and addition of new phases such as P_{diff} indicates that *P*-wave speed models may be more reliable than previously thought (Houser et al., 2008; Manners, 2008). An example of a CMB heat flux map produced in this manner from a *P*-wave speed tomography model is shown in **Figure 21** after assuming a maximum heat flux of 160 mW m^{-2}.

One quantity that can be estimated from these approaches is the relationship between the maximum CMB heat flux q_{max}, and heat flow Q_{cmb}. Any normalized heat flux model can be integrated to yield the proportionality:

$$Q_{\mathrm{cmb}} = A^* q_{\mathrm{max}} \qquad [13]$$

The coefficient A reflects the distribution of heat fluxes obtained from the seismic tomography model, and is unique to every model. For example, in the simple model shown in **Figure 21**, one finds that $A = 81$ (with Q_{cmb} in units of TW and q_{max} in units of W m^{-2}) so that a maximum heat flux of 160 mW m^{-2} would yield a total CMB heat flow of 13 TW,

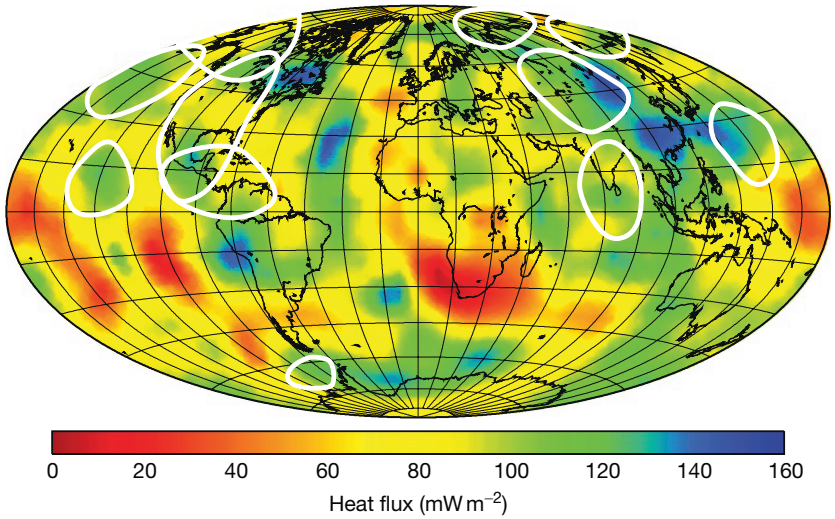


Figure 21 CMB heat flux map produced by assuming a proportionality between temperature and *P*-wave speed velocities using the Scripps *P* model (Houser et al., 2008). In this example, a maximum heat flux of 160 mW m^{-2} is used, corresponding to 13 TW global CMB heat flow. Notice the blue regions of highest heat flux, which can skew the distribution and may be due to features other than heat flux anomalies. Observations of *D*″ discontinuities and/or strong seismic anisotropy from **Figure 4** are outlined by thick white lines. In terms of the PPv-double-crossing model described earlier, PPv lenses are predicted to occur in regions with greater than ~80 mW m^{-2} heat flux. This crossover heat flux corresponds to a thermal gradient of ~8 K km^{-1} for a thermal conductivity of 10 W m^{-1} K^{-1}, consistent with Pv–PPv phase boundary gradients of 4.2–6.1 K km^{-1} (for Γ = 9–13 MPa K^{-1}).

consistent with constraints from plume heat flow as discussed previously (and also consistent with core heat flow, as will be discussed in the succeeding text). If one performs the same procedures using a corresponding *S* model (Houser et al., 2008), one obtains $A = 100$; thus, even with the complexities in using *S* discussed earlier, we arrive at a similar result. In general, most models yield similar answers because the dominant distribution of wave speed variations is approximately Gaussian in form (Hernlund and Houser, 2008); thus, the normalized variations have the a similar form and differences in *A* between models are probably more sensitive to the extreme tails of the distribution.

Employing a proportionality between heat flux and seismic-wave speeds, as discussed in the preceding text, is roughly equivalent to assuming that the boundary layer thickness is everywhere the same and that variations in heat flux are mainly due to variations in temperature mapped from seismic-wave speed. However, the boundary layer is expected to be thicker in hotter regions and thinner in colder regions; this effect should amplify differences in implied heat flux for fast and slow regions in seismic tomographic models relative to a simple proportionality. Wu et al. (2011) integrated the heat flux variations using a more dynamically consistent boundary layer model to yield a total heat flow of 13 ± 3 TW, which is probably the most accurate application among all previously proposed techniques for extrapolating heat flux to the entire CMB. Yet, despite the addition of further complexities, the result is similar to prior estimates, converging to ~13 TW.

7.11.7.2.6 Summary of D″ constraints upon CMB heat flow

The discovery of the Pv–PPv phase change and improvements in seismic tomography have revolutionized approaches for estimating CMB heat flow from *D*″ structure and have moved beyond rough estimates based on temperature change and boundary layer thickness. While there are still some residual uncertainties in estimating thermal gradients in the deepest mantle, the variation in permissible values is probably within a factor of 2, which is a significant improvement. The convergence of estimates for the lowermost mantle thermal conductivity around ~10 W m^{-1} K^{-1} is also promising, although as noted previously, there may still be some uncertainties, and further research may yet give rise to revisions. The most carefully developed constraints for CMB heat flow (Wu et al., 2011), which rely in part on inferences from the double-crossing model, yield values of CMB heat flow in the range 13 ± 3 TW. CMB heat flow of 13 TW corresponds to a maximum CMB heat flux of around 160 mW m^{-2}, which is more than sufficient to account for a PPv-double-crossing.

7.11.7.3 Heat Flow Constrained by Core Evolution and Dynamics

The most recent major development in core–mantle heat exchange constraints comes from upward revisions in estimates for the conductivity of the Earth's core. Constraints such as plume heat flow and *D*″ thermal boundary layer considerations discussed previously can only constrain the Earth's present CMB heat flow, but do not necessarily constrain CMB heat flow in the distant past. Inferences from mantle circulation models depend on plate boundary reconstructions, which themselves are limited to the age of the present seafloor (i.e., several hundred Ma). Constraints involving core conductivity, on the other hand, extend into the Earth's ancient past because they are anchored by the earliest paleomagnetic observations of a geomagnetic field at ~3.5 Ga (Tarduno et al., 2010). Furthermore, while other constraints have some residual

uncertainties and ongoing debates regarding their magnitude and applicability, the recent research on core conductivity, coming from a variety of different approaches, has converged on a surprisingly consistent result. In the following, we give a brief overview of this recent development.

7.11.7.3.1 Heat carried by core conduction and convection

Heat transport in the core differs from that in the mantle in many respects. Sufficiently vigorous convection causes a medium to trend toward a well-mixed isentropic state. Convected heat is carried in the core via radial transport of material exhibiting fluctuations in temperature relative the isentropic value. Even though lateral temperature variations in the core are less than about 10^{-4} K, when coupled with radial flows of order 0.1 mm s^{-1} – consistent with observations of geomagnetic secular variation – it is easy to produce convected heat flows of order ~1 TW (e.g., using $\sim pC_p v_z \Delta T$). Unlike the mantle, the relatively high conductivity and isentropic gradient in core material yield a very large conducted heat flux, which was previously thought by many researchers to be TW at the CMB (e.g., Buffett, 2002; Gubbins et al., 2003; Labrosse et al., 1997) but has recently been upwardly revised to values exceeding 10 TW (de Koker et al., 2012; Gomi et al., 2013; Pozzo et al., 2012). The sum of conducted and convected heat out of the core should be equivalent to the CMB heat flow, and therefore, heat conduction is probably greater than the heat carried by core convection.

There exist a variety of useful equivalent expressions for the isentropic gradient, such as

$$-\left(\frac{\mathrm{d}T}{\mathrm{d}r}\right)_S = \frac{T\alpha g}{c_p} = \frac{T\gamma g}{\phi} = T\gamma\left(\frac{\partial \log \rho}{\partial r}\right)_S. \qquad [14]$$

where T is temperature, r is radius, S is entropy, α is the thermal expansivity, γ is the Grüneisen parameter, and ϕ is the seismic parameter, which is defined as K_S/ρ (where K_S is the isentropic bulk modulus), or the square of the sound speed in the liquid outer core. Quantities such as ϕ, $(\partial \log \rho/\partial r)_S$, and g are very accurately determined by seismological models that fit a large volume of seismic data to the adiabatic compression trend for the entire outer core (e.g., PREM, Dziewonski and Anderson, 1981), such that they likely contribute less than 1% error to the determination of the large-scale isentropic gradient. Greater uncertainties in estimating $(\mathrm{d}T/\mathrm{d}r)_S$ arise from the parameter combinations $T\alpha/c_p$ or $T\gamma$. Constraints derived from the PPv double-crossing suggest CMB T to be around 3750 K (Tateno et al., 2009), consistent with an inner-core boundary T of ≈5000 K, although uncertainties at both boundaries probably exceed 10%. The ratio α/c_p not only can be estimated using an experimentally constrained thermal equation of state for iron (e.g., Dewaele et al., 2006) but also likely carry residual uncertainties of at least 10%. While ab initio models should be able to calculate quantities such as γ with relatively high precision, yielding values close to $\gamma = 1.5$ (see Vočadlo et al. (2003) for a review), the errors in this parameter could also be ~10%. Assuming a CMB temperature of 3750 K, with the Dewaele equation of state, one obtains $(\mathrm{d}T/\mathrm{d}r)_S = 0.89$ K km^{-1}, while using $\gamma = 1.5$ and PREM yields 0.91 K km^{-1}. Thus, these semi-independent estimates yield very similar values, although uncertainties of order 10% still remain. In the following, we will adopt the value 0.9 K km^{-1} for the purpose of estimating conducted heat flux.

The primary uncertainty in estimating the heat conducted along the isentropic gradient has always been the thermal conductivity of impure iron–nickel-alloyed liquid metal at the P–T conditions of the core (i.e., $P > 136$ GPa, $T > 4000$ K). In metals, the thermal conductivity is dominated by transport of heat by electrons, with lattice contributions accounting for ~10% of the total conductivity. The electronic part of the conductivity, k_e, is linearly related to the electrical conductivity σ by the Wiedemann–Franz law:

$$k_e = L_0 \sigma T, \qquad [15]$$

where $L_0 \approx 2.45 \times 10^{-8}$ WΩ K^{-2} is the Lorenz constant. Thus, the dominant contribution to thermal conductivity can be estimated from electrical conductivity from resistivity constraints. However, the measurements cannot be robustly extrapolated across phase transitions and must be performed at very high pressures in the phase stability region of interest.

The first inferences of thermal conductivity at core pressure were derived from shockwave measurements, which yielded values in excess of 100 W m^{-1} K^{-1} for both pure iron (Bi et al., 2002; Keeler and Mitchell, 1969) and iron–silicon alloy (Matassov, 1977). Given concerns regarding the relatively high CMB heat flows implied by these estimates, they were not taken as seriously as lower estimates. Stacey and Anderson (2001) downwardly revised the estimate of k_e to 42 W m^{-1} K^{-1}, and Stacey and Loper (2007) later suggested even lower values for k_e around 30 W m^{-1} K^{-1}, permitting thermal evolution scenarios with low CMB heat flow. More recently, *ab initio* modeling techniques have been developed to study core conductivity and have unanimously found that core conductivity is closer to 100 W m^{-1} K^{-1} atop the core, more compatible with the earlier inferences based upon shock measurements (de Koker et al., 2012; Pozzo et al., 2012). High-pressure ambient-temperature DAC conductivity measurements have exhibited excellent agreement with ab initio calculations on iron and iron–silicon alloys at core pressures, building confidence in the computational model results, although they were only able to directly compare at ambient temperatures owing to experimental limitations (Gomi et al., 2013). However, Gomi et al. (2013) demonstrated that all of the high P–T ab initio results, high-P and ambient-T DAC experimental results, and even the older shockwave measurements are in good agreement with each other if one adopts a resistivity saturation model, which has been well known in the materials science literature for half a century (e.g., Gunnarsson et al., 2003) but was never applied in the geophysical context until now. In fact, saturation resistivity provides a simple theoretical basis for interpreting all of the higher conductivity values derived from every available technique and establishes an absolute theoretical minimum of $k_e \approx 60$ W m^{-1} K^{-1} at the top of the core (Gomi et al., 2013). In summary, very rapid progress in understanding core conductivity has now established estimates for thermal conductivity of $k_e > 90$ W m^{-1} K^{-1} at the CMB, increasing even further with depth in the core below the CMB.

The conducted heat flux implied by a 0.9 K km^{-1} isentropic gradient and thermal conductivity in excess of 90 W m^{-1} K^{-1} is greater than 77 mW m^{-2}, producing a CMB conducted heat

flow greater than 11.6 TW, very close to the values for the CMB heat flow estimated from the mantle constraints discussed previously. Another interesting implication of this result is that the inner core would have begun to crystallize about 0.5 Ga (Gomi et al., 2013) and is therefore only a relatively recent feature of the Earth (approximately Cambrian). This changes very little even if one considers the possible presence of a plausible amount of radioactive potassium in the core (Labrosse, 2003). In the absence of an alternate mechanism for producing the geodynamo, convection in the core prior to the inner core would have been driven solely by thermal buoyancy, requiring a positive convected heat flow in addition to the conducted heat flow. If the heat flow falls below the amount conducted down the adiabatic gradient, then the core would begin to thermally stratify and convection would become impeded. It is unknown precisely how much stratification the Earth's core could tolerate while still producing a geomagnetic field at the surface compatible with paleomagnetic constraints and is a topic that needs to be further investigated. Furthermore, given the fluctuations in CMB heat flow expected over time owing to processes such as periodic assembly and dispersion of continents (i.e., the 'Wilson cycle'), the average CMB heat flow would need to be significantly higher than the lower threshold in order that lower heat flow periods would not shut off the geodynamo (Jaupart et al., 2007). Thus, the conducted heat flow of 11.6 TW likely represents a lower bound on average CMB heat flow during the period 0.5–3.5 Ga prior to the inner core, during the era that the Earth is known to have produced a magnetic field consistent with a geodynamo (Tarduno et al., 2010). Additionally, the conducted heat flow was somewhat higher in the past given the influence of hotter conditions on both the isentropic gradient and thermal conductivity (Gomi et al., 2013).

7.11.7.4 High CMB Heat Flow and Thermal Evolution

Recent estimates for CMB heat flow converge to values of ~13 TW, including plume heat flow, estimates of temperature gradients and thermal conductivity in D'', and a high conducted heat flow down the core isentrope. The possibility for such high heat flow has raised concerns regarding the implied secular cooling of the core and deep mantle. The core has a mass of 2.1×10^{24} kg (Dziewonski and Anderson, 1981) and specific heat of 740 J K^{-1} kg^{-1} (Vočadlo et al., 2003), yielding a total heat capacity of about 1.6×10^{27} J kg^{-1}. Taken over the age of the Earth prior to the formation of an inner core (i.e., before 0.5 Ga), 13 TW heat flow translates to 1.6×10^{30} J, corresponding to ~1000 K average cooling. (For comparison, this is roughly an order of magnitude larger than the rotational kinetic energy of the present-day Earth.) The core has cooled by ~100 K since the birth of the inner core (e.g., Labrosse, 2003); thus, the total secular cooling for 13 TW CMB heat flow is expected to be 1100 K. Keep in mind, however, that this is a lower bound since heat flow is variable and would have been higher, on average, in the past.

There are two major problems with such a large secular cooling of the core. The first is that it is much greater than the secular cooling of the mantle inferred from igneous petrology constraints (e.g., Herzberg et al., 2010), and even with the possibility of chemical layering in the deep mantle, it may be difficult to shield the shallow mantle from plumes rising up from the much hotter deep mantle (e.g., Farnetani, 1997). The second problem is that the temperature of the CMB during the early stages of the Earth's evolution would have been significantly hotter than the liquidus for any reasonable bulk composition (e.g., Andrault et al., 2011; Boehler, 2000; Fiquet et al., 2010; Zerr et al., 1998), and the lowermost mantle should therefore have been molten for an extended period of geologic time (Labrosse et al., 2007). It has been proposed that such a melt would have to be dense enough to remain stable at the bottom of the mantle for the several Gy periods of time prior to temperatures falling to the mantle solidus, since a buoyant melt would otherwise have destabilized the boundary layer, leading to rapid secular cooling to the solidus in the very early Earth, which is incompatible with sustained CMB heat flow of 13 TW or higher. The presence of some kind of long-lived dense BMO is therefore an inescapable outcome of the convergence of CMB heat flow constraints around relatively high values. However, it is not clear that the BMO model can account for such high temperatures in the early Earth, since the implied CMB temperature may exceed the liquidus by an amount greater than predicted in simple models (Labrosse et al., 2007).

7.11.8 Core–Mantle Mass Exchange

It is simplest to think of the Earth's core as a chemically isolated system, having a composition that reflects processes that predate the cessation of the Earth's formation, such as silicate–metal equilibration in a deep magma ocean during the Earth's formation (e.g., Li and Agee, 1996; Wood et al., 2006), and with negligible exchange of matter in the subsequent ~4.4 billion years. On the other hand, it has been hypothesized that some isotopic and chemical signatures may originate from core–mantle mass exchange, such as an elevated Fe/Mn observed in OIBs (Humayun et al., 2004) or Os isotopic fractionation reflecting inner-core growth (Brandon and Walker, 2005). Indeed, recent CMB thermal evolution models (Labrosse et al., 2007), seismic observations (Rost et al., 2005), and inferences from high-P–T experiments (Walker et al., 2002) together highlight the CMB as a potentially highly chemically reactive zone reflecting a large degree of disequilibrium between the average core and mantle at CMB P–T conditions. Understanding the degree of core–mantle mass exchange is important for understanding the long-term evolution of the CMB region, as well as the origin and evolution of its major features.

7.11.8.1 Core Versus Mantle Affinity

Before discussing mass transfer between the mantle and core, it is useful to make a distinction between the material identified as having an affinity for the core and the material that has properties belonging to the mantle. Essential properties of the core matter include a high density (closer to 9900 kg m^{-3}), a high enrichment in iron (e.g., Fe mole fraction $\geq 1/2$), and a metallic character. Essential properties of the mantle matter include a smaller density (closer to 5.5 kg m^{-3}), a high alkali content (i.e., Ca, Na, and Mg), and an ionic character. Some

species, such as O and Si, are equally soluble in the core or mantle matter at CMB conditions and therefore are of little use as a discriminant. Notice that we also avoid using solid versus liquid as a discriminant, since it is not diagnostic of the difference between the core and mantle (both of which contain solids and liquids). The reason for the necessity of establishing this kind of definition is that it is possible for mantle matter to be converted into core matter, and vice versa, such that the properties of 'mantle-ness' and 'core-ness' are not strictly conserved. In this more general view, the CMB is a boundary that may shift through material by transformation of the matter into one form or the other.

There are two distinct ways to transfer material between the mantle and core: (1) by mass transport in which core and mantle materials move relative to one another in opposite directions perpendicular to the CMB and (2) by conversion of the mantle matter into the core matter, or vice versa. In the first case, material of one affinity can be said to be dynamically 'entrained' into the region of matter having the other affinity, but without losing its essential identity. In the second case, the identity of a parcel of matter itself is changed from one kind to the other. Both kinds of mechanisms for core–mantle mass transfer have been proposed in the past.

7.11.8.2 Entrainment and Other Direct Core–Mantle Exchange Processes

Once material has crossed the CMB, the question becomes whether or how it will be transmitted or mixed into the bulk of the mantle or core. However, it is crucially important to recognize that core–mantle mass exchange may be confined – to a greater or lesser degree – to the deepest mantle and/or the topmost outer core by hindrances to radial transport. For example, matter entering the mantle from the core may tend to be inherently dense and exhibit a relatively low viscosity (e.g., due to Fe enrichment), both of which can hinder or slow the viscous entrainment of this material upward to shallower depths. Likewise, matter transmitted from the mantle to the core could be less dense than the underlying core, and downward entrainment in this context is very difficult owing to the virtual absence of viscous forces in the outer-core fluid (although turbulent entrainment might be possible). Still, some degree of entrainment and transport away from the CMB is expected in all cases, though the flux may be limited to a greater or lesser degree depending on a variety of complex factors (e.g., Buffett and Seagle, 2010; Sleep, 1988).

Several kinds of entrainment may occur at the CMB; however, most of the proposed mechanisms involve intrusion of liquid core metal into the base of the mantle rather than the mantle into the core. The latter case is probably ruled out by the very brief Stokes rise time for buoyant mantle material in the core, in addition to the weak ability for core flows to mobilize and/or entrain particles (Solomatov et al., 1993). Percolation of fluid through a solid matrix relies on the ability for the liquid to form an interconnected grain-wetting network within solid crystalline rock via decompaction (von Bargen and Waff, 1986). While the strong surface tension of metals can inhibit wetting, this may be less effective at high temperatures and impurity concentrations. Some experiments at conditions close to CMB P–T support grain wetting between iron and PPv-bearing rocks for liquid metal fractions greater than ~2% (Sakai et al., 2006). Thus, iron may be likely to penetrate into, and flow through, mantle rocks at CMB conditions by decompaction and percolation. However, the strong buoyancy force between metals and silicates/oxides at the CMB acts to drain fluid back into the core. It may be possible to support a limited degree of intrusion by capillary forces; however, this is estimated to accommodate only limited intrusion of order 100 m or less (Poirier et al., 1998). It has also been suggested that poroviscoelastic entrainment of core fluid can occur (Petford et al., 2005); however, it is not clear that the requisite dilatant elastic properties exist for materials at D'' conditions. Kanda and Stevenson (2006) proposed a simple mechanism for obtaining core fluid intrusion into the mantle, which relies upon decompaction and percolation but is driven by downward deflection of the CMB into the core, such as the dynamical topography expected beneath the mantle downwelling currents. In this case, the limiting depth extent of intrusion will be similar to the dynamical topography itself, of order 1 km (e.g., Olson et al., 1987; Tanaka, 2010). Once the mantle material moves away from regions of downward deflection into the core, it will rise upward again and much of the core fluid could be expected to drain back into the core. However, a small residual fraction of order 1% may not completely wet grain boundaries and could therefore become frozen into the rock matrix. Entrainment of iron into mantle rocks, even at small fractions, will affect the bulk density of the rock, and might cause it to become dense enough to resist upward transport into the overlying mantle. However, as discussed previously, in the regime of viscous entrainment that dominates mantle dynamics, such material will gradually be eroded away and mixed into the overlying mantle.

7.11.8.3 Core–Mantle Chemical Reactions

CMB chemical reactions involve the exchange of chemical components between phases belonging to the core and mantle and occur at the atomic scale by interdiffusion of molecular components across the interface between these phases. These chemical reactions occur whenever the *chemical potentials* (μ) of mutually intersoluble components are unequal; thus, CMB disequilibrium is manifested as a discontinuity or gradient in chemical potential(s) across the CMB. (Note: a discontinuity or gradient in *concentration* does *not* imply disequilibrium.) Core–mantle disequilibrium may be an important driving force for the long-term evolution of heterogeneous structures at the CMB, and is probably relevant to ULVZ as the veneer separating the overlying mantle from the top of the core. The top of the core may be chemically distinct from the deeper core, particularly if the stratification inferred by seismology exists, as discussed in Section 6.

It is helpful to clarify the meaning of terminology used in this context. For example, the *solubility* of a given component in a particular phase (e.g., oxygen in iron liquid) is not always a clearly defined concept. In some cases, there may be a *solubility limit* for a given solute in a specific phase, and a concentration at this limit is said to be at *saturation* or to be *saturated*. Concentrations below or above any such limit are *undersaturated* or *oversaturated*, respectively. However, in some cases, there is no solubility limit, and a mixture of 2 components in a phase can

be said to exhibit complete solution series between the pure end-members, in which case they may also be said to be *miscible*. Yet, even if a solubility limit does occur, saturation is not always relevant to the actual concentration(s) of solute(s) expected in a phase (i.e., solvent). In many cases, we are more interested in the *equilibrium concentration*, which is always less than the saturation limit (if one exists) and is instead determined by chemical equilibrium between **2** or more phases (solvents) which are able to freely exchange the component (solute) of interest. A component i, which is soluble in both phase j and phase k, is said to be in chemical equilibrium whenever $\mu_i^j = \mu_i^k$, where μ is the chemical potential and is a function of the bulk chemical composition of the phase in addition to the P–T conditions. This relationship is as fundamental as equality of temperature between two bodies at thermal equilibrium.

In the equilibrium partitioning case, the concentration of a component in one phase also depends on the concentration of that component in the other phase(s), and these concentrations are generally distinct from one another. In the case of a saturated solution, on the other hand, the concentration of the component in the saturated phase does not depend on the concentration of that component in other phases, since it is at the limit of its solubility. Being saturated in a component does not, however, preclude the relevance of equilibrium exchange of that component at subsaturated concentrations. A phase with concentrations of a component above the saturation limit is metastable and said to be *supersaturated*; such a mixture is prone to precipitation of another phase that hosts the excess solute. Supersaturation can occur, for example, as a saturated (or undersaturated) phase cools or rises to lower pressures, in which case the solubility might drop below the existing concentration (as happens in degassing of magmas during ascent). Core–mantle mass exchange processes driven by saturation and/or equilibrium between phases have been proposed in the past, and we discuss various examples in the following.

7.11.8.3.1 Length scales of equilibration

Although there exists large-scale disequilibrium between the core and mantle (discussed in the next section), at sufficiently small scales, it is possible to achieve 'local equilibrium,' although one needs to be careful about how this is defined. If a small parcel of core fluid of length scale L_c is brought into contact with a small parcel of mantle of length scale L_m under conditions of chemical disequilibrium, the exchange of components between them by diffusion offers an opportunity for these to approach equilibrium since net diffusion rates are driven by differences in chemical potential. The timescale for equilibration in such a case is thus limited by the diffusion timescales $\tau_c = L_c^2/D_c$ and $\tau_m = L_m^2/D_m$, where D_c and D_m are the limiting (i.e., slowest) diffusivities of interacting components in the core and mantle matters, respectively. Recognizing a large degree of uncertainty in estimates of the limiting diffusivities, we can obtain rough order of magnitude estimates using values of $D_c \sim 10^{-8}\ m^2\ s^{-1}$ (Vočadlo et al., 2003) and a much larger range of potential variation for relevant mantle materials $D_m \sim 10^{-16}\ m^2\ s^{-1}$–$10^{-14}\ m^2\ s^{-1}$ (Van Orman and Crispin, 2010).

There is one further caveat for local equilibrium that is important to keep in mind. In order to approach equilibrium, the diffusion timescale of locally equilibrating matter must be smaller than the timescales of other processes that are perturbing the equilibrium, such as secular changes in temperature or composition as discussed previously. The timescale for secular shifts in equilibrium at the CMB is typically very slow over geologic timescales (i.e., of order 10^8 years), although it still presents a limit upon the size of regions in the core or mantle that may approach equilibrium.

For a core parcel of scale L_c and a mantle parcel of scale L_m, equating the diffusion timescales gives

$$\frac{L_c}{L_m} = \sqrt{\frac{D_c}{D_m}}. \qquad [16]$$

In words, the size of a core parcel that may equilibrate with a mantle parcel on a given timescale is larger by a factor $\sqrt{D_c/D_m}$. For processes relevant to the Earth's CMB, $\sqrt{D_c/D_m} \sim 10^3 - 10^4$. Therefore, 'local equilibrium' can be approached over significantly larger length scales in the core than in the mantle. Assuming a typical residence time for mantle material convecting along the CMB of 10^8 years, the thickness of the zone that is able to approach equilibrium on the mantle side is only of order 0.1–10 m. Over the history of the Earth, material with this residence time would have been circulated through the CMB region ~45 times, accounting for a total layer thickness of order 10–1000 m at the CMB. Even if one were to account for variable rates of mantle convection through geologic history, core–mantle reactions limited by diffusion length scales alone would only affect a very small fraction of mantle (about 10–100 ppm by volume). The volume of core that would be affected could be much larger, although the influence of reactions on core dynamics may be more severe (e.g., leading to stratification, Buffett and Seagle, 2010).

It is possible to dramatically increase the length scales of mantle reaction zones if core fluid intrudes into the mantle along grain boundaries by decompaction, as discussed previously. It is still required that the metals and oxides are mixed/homogenized to a scale finer than 0.1–10 m (see the preceding text) in order for equilibrium to be approached; thus, in the slowest diffusion case, it may be necessary to mix to length scales approaching the grain size, which is readily satisfied by grain-scale percolation mechanisms. For mantle flow through the CMB region with an average residence time of about 100 My, the dynamical topography-driven mechanism proposed by Kanda and Stevenson (2006) would accommodate interaction with a volume of mantle equivalent to ~50 km thickness at the CMB over the Earth's history. On the other hand, the volume of core matter available for reaction may be significantly smaller if metals become largely isolated from the top of the core by virtue of a complex intergrain geometry and/or stabilizing composition-induced density gradient owing to reactions (e.g., by enrichment of metals in oxygen). Thus, a more significant volume of mantle may be reacted with the core if metal intrudes by decompaction, and as mentioned earlier, some residue of this metal could be trapped inside the rock matrix below a wetting fraction of order 1%. Core fluid reacted with the mantle in this manner will drain back into the

core; however, if it contains even a very small enrichment in light alloys (Si, O, etc.), it may be impossible for the core to downwardly entrain the fluid and mix it into the deeper core.

7.11.8.3.2 Disequilibrium-driven reactions

It is important to recognize that, even in the mildest of circumstances, the CMB should be a chemically reactive zone for much of the Earth's history (Stevenson, 1981). Core–mantle disequilibrium may be an inherent consequence of the manner in which the Earth's core formed and is constantly perturbed by secular changes during the Earth's evolution since formation. It is thought that elemental abundances in the silicate Earth reflect metal–silicate equilibrium during a shallower magma ocean phase in differentiating planetesimals or the proto-Earth (e.g., Wood et al., 2006), which are distinct from the much higher-pressure and higher-temperature conditions that have prevailed at the CMB since formation. Therefore, the average core and mantle should be far from equilibrium at CMB P–T from the very beginning of the Earth's history. Furthermore, equilibrium is generally temperature sensitive, and the CMB undergoes secular cooling over time (perhaps more than 1000 K, as discussed previously). Thus, secular cooling continuously shifts the bulk of the core and mantle away from any equilibrium that may have been established in the past. Finally, secular changes in the composition of the core and/or mantle materials circulating through (or accumulating at) the CMB are expected, causing another perturbation to the disequilibrium between the core and mantle in the CMB region itself. For example, such secular composition changes can arise in the mantle as a result of the accumulation of MORB (e.g., Christensen and Hofmann, 1994) or banded-iron formation (Dobson and Brodholt, 2005) or from fractional crystallization of a BMO (Labrosse et al., 2007), while in the outer core, a secular enrichment in incompatible alloys is expected as the inner core grows by fractional crystallization (Braginsky, 1963).

On the basis of experiments performed up to mid-mantle P–T conditions in the DAC, it has long been suspected that some silicates and oxides may readily dissolve into metallic iron at high P–T (e.g., Dubrovinsky et al., 2004; Knittle and Jeanloz, 1991). Knittle and Jeanloz (1991) combined (Mg, Fe)SiO_3^{Pv} and Fe^M (M = metal) at high P–T in a laser-heated DAC and, upon observing Fe-free Pv, along with FeO and FeSi in the quench products, suggested the reaction

$$(Mg, Fe)SiO_3^{Pv} + Fe^M \rightarrow FeO^M + FeSi^M + MgSiO_3^{Pv} \qquad [17]$$

In the proposed reaction, the $FeSiO_3$ component of Pv decomposes into FeO^M and $FeSi^M$. Such reactions would not be limited to Pv; other oxide phases such as ferropericlase or aluminous phases may also exhibit reactive tendencies with the core (e.g., Dubrovinsky et al., 2004). Some have speculated that FeO^M and $FeSi^M$ reaction products could accumulate in the CMB region as solid phases (Buffett et al., 2000; Kellogg and King, 1993; Manga and Jeanloz, 1996); however, the melting temperature of either species is ~4000 K or less (Fischer and Campbell, 2010; Lord et al., 2010), very similar to estimates of the present CMB temperature. The CMB was much hotter in the first 3 Gy of the Earth's history, ensuring that FeO and/or FeSi would have been molten at the CMB in the past, possibly only reaching the liquidus temperature in recent geologic history. Indeed, recent studies (e.g., Asahara et al., 2007; Frost et al., 2010; Komabayashi, 2014; Ozawa et al., 2009; Tsuno et al., 2013) view core–mantle reactions in the Earth as a coupled dissolution of O (and possibly Si) into the liquid metal core as metallic alloying species, instead of forming stoichiometric solid phases as products.

The Knittle–Jeanloz reaction suggests that the behavior of oxygen exerts an important influence upon CMB reactions. Consider simple dissolution of ferric iron oxide from silicate (*Sil*) into liquid metal (*M*) at equilibrium:

$$FeO^{Sil} \rightleftharpoons O^M + Fe^M. \qquad [18]$$

This equilibrium follows a relation involving a Gibbs free energy change for the reaction ΔG_r according to,

$$K^{M/Sil} = \frac{a_{Fe}^M a_O^M}{a_{FeO}^{Sil}} = \left(\frac{\gamma_{Fe}^M \gamma_O^M}{\gamma_{FeO}^{Sil}}\right)\left(\frac{X_{Fe}^M X_O^M}{X_{FeO}^{Sil}}\right) = \exp\left(\frac{\Delta G_r}{RT}\right), \qquad [19]$$

where K is the (metal–silicate) equilibrium constant, a is the activity of a component (subscript) in a given phase (superscript), γ is the activity coefficient of a component in the designated phase, X is the mole fraction of component in the phase, and R is the gas constant. In the ideal case, at a given pressure, the activity coefficients are unity and the temperature dependence follows the usual Arrhenius relation, with higher temperatures favoring the higher entropy side of the equilibrium (in this case, the metal components). The pressure dependence of the equilibrium, on the other hand, is governed by volume changes of species from the left side to the right side. For example, if the partial molar volume of oxygen in the metal is relatively large in comparison with its partial molar volume in oxides, then reactions involving incorporation of oxygen into the metal will be impeded at elevated pressures. Past debates focused on the possibility of either a diminishing oxygen solubility in metal with increasing pressure (e.g., Rubie et al., 2004) or only a modest change with pressure (e.g., Walker et al., 2002). Some of the early work on this system was performed at relatively low pressures and was not representative of the behavior of the high pressure (hcp) form of iron. However, recent experiments at greater pressures and temperatures (Asahara et al., 2007; Ozawa et al., 2008, 2009; Tsuno et al., 2013) and thermodynamic models (Frost et al., 2010; Komabayashi, 2014) typically yield little or no discernible pressure dependence for the reaction [18]. Thus, the emerging picture is one in which oxygen solubility is only weakly dependent upon pressure but exhibits at least an Arrhenius dependence in temperature. The activity coefficients relevant to this reaction are not well constrained at core conditions, although recent experimental evidence argues for approximately ideal behavior (Komabayashi, 2014). Note that these coefficients can vary with temperature, pressure, and composition; thus, a thorough description of equilibrium is somewhat complex.

The coupled behavior of silicon and oxygen at CMB conditions is also an important consideration, since both species are not mutually soluble in liquid iron at ambient pressures. One way of describing coupled Si and O behavior is to invoke a Si/O exchange equilibrium:

$$SiO_2^{Sil} + 2Fe^M \rightleftharpoons Si^M + 2FeO^{Sil} \qquad [20]$$

Previous experimental work has been divided, with some supporting the incorporation of oxygen and silicon

together Takafuji et al. (2005) and others indicating a limited mutual solubility of oxygen and silicon in metal (e.g., Mann et al., 2009; Ricolleau et al., 2011). A recent model has been proposed that yields a significant enhancement in coupled Si + O solubility in metals at temperatures above ~3000 K (Tsuno et al., 2013). Other recent models consider equilibrium between the volumetrically dominant lower-mantle phases $(Mg,Fe)SiO_3^{Pv}$ and $(Mg,Fe)O^{Pe}$ (Pe = periclase) and (liquid metal) iron, for which the equilibrium can be expressed as

$$FeO^{Pe} \rightleftharpoons Fe^{M} + O^{M}, \quad [21]$$

$$FeSiO_3^{Pv} \rightleftharpoons Fe^{M} + Si^{M} + 3O^{M}, \quad [22]$$

and for closure, one must also add another expression for the iron exchange between Pe and Pv:

$$FeSiO_3^{Pv} + MgO^{Pe} \rightleftharpoons MgSiO_3^{Pv} + FeO^{Pe}. \quad [23]$$

When calibrated with experimental data, this type of model predicts a strong tendency for oxygen to dissolve into the metal at CMB conditions and for typical mantle iron concentrations (e.g., Ozawa et al., 2009). While Si also dissolves into the metal, its tendency to dissolve into the metals is relatively modest. Thus, different models and experiments yield a variety of behaviors, and it is hoped that further work at higher temperatures and pressures will lead to better agreement.

One of the most interesting features of metal–silicate equilibrium models is that an increase in FeO in the silicates is coupled with an increase in oxygen dissolved in metal. For example, in eqn [19], the equilibrium constant is invariant if a_{FeO}^{Sil} increases in proportion to a_O^M. Thus, a higher oxygen content in the metal is associated with a higher iron fraction in the silicate, all else being equal. The recent models and experiments for equilibrium of FeO-bearing components in mantle phases with O dissolved in the core (Asahara et al., 2007; Frost et al., 2010; Ozawa et al., 2008, 2009) suggest that the amount of oxygen that would be present atop the core in equilibrium with typical mantle compositions (e.g., pyrolite, with $Mg\,Mg^{-1} + Fe \approx 90\%$) would exceed the 10 wt% total budget of impurity species in the isentropic convecting region of the outer core by a significant amount (i.e., by several tens of percent). The amount of oxygen in the core liquid is even larger if the base of the mantle is Fe-enriched, consistent with seismic inferences of ULVZ. Thus, these experiments initially seemed to present a kind of paradox, in that the mantle should either be depleted in iron or the core should contain more oxygen. Neither possibility is consistent with seismological constraints.

Asahara et al. (2007) proposed that the base of the mantle is depleted in iron, allowing the equilibrium partitioning to be compatible with the seismologically inferred density of the outer core. However, these models present problems for explaining seismic observations of very dense Fe-rich structures at the CMB (e.g., ULVZ as discussed previously). The models additionally assume that oxygen and/or silicon dissolved into the top of the core following such a reaction would be readily mixed downward into the deeper interior of the outer core. Even if this were to occur, such a process would increase the GPE of the core, thus impeding convection and generation of a geodynamo. For these reasons, a model with a Fe-depleted mantle side in equilibrium with the entire convecting outer core is not satisfactory.

As discussed previously, Buffett and Seagle (2010) proposed an alternative model in which dissolution of oxygen into the top of the core leads to the growth of a buoyant gravitationally stratified layer, since it is likely to be difficult for flows in the outer core to entrain material with a density anomaly of order 1%. In this scenario, the stratified layer encroaches downward into the top of the core from the CMB at a rate controlled by diffusion of oxygen, yielding a layer ~100 km thick over the age of the Earth. The very bottom of the layer could be partly entrained into core flows, presenting an energetic drag upon core convection, but this effect is more limited than the case where all reaction products become entrained into the convecting outer core and is self-regulated by the convective power available for driving convection. Also, since the top of the core is oxygen-enriched, the very bottom of the mantle need not be as depleted in iron. However, the base of the mantle is still depleted in FeO-bearing components in this model, leading to contradictions with seismological observations of very dense (likely Fe-rich) structures at the CMB.

One possibility to resolve all of these issues, while satisfying seismological observations, is that the very bottom of the mantle has been gradually enriched in FeO by processes such as fractional crystallization of liquid over geologic time, as proposed in the BMO model of Labrosse et al. (2007). This would increase the FeO content at the base of the mantle while also increasing the oxygen content at the top of the core. One would still arrive at an oxygen-enriched stratified layer atop the core similar to the Buffett and Seagle model, but in this case, the very base of the mantle would be left Fe-enriched, while the top of the core would contain relatively more oxygen (hence more buoyant and stable).

7.11.8.3.3 Core saturation?

Another possible means of transferring a significant amount of mass between the core and mantle is exsolution, in which a saturated species rises up to the CMB and ostensibly becomes part of the mantle. Precipitation of FeSi and/or FeO solid crystals from the outer core was suggested by Buffett et al. (2000); however, such a mechanism depends on saturation of the core fluid in Si and/or O, which may be unlikely if both are mutually soluble at high temperatures. Another kind of saturation mechanism is often suggested by Stevenson (2012). If some of the metal that formed the Earth's core equilibrated with oxides at very high temperatures during the Earth's formation, then they could possibly have ingested components such as magnesium, which would otherwise be very incompatible in the metal phase. With progressive core cooling, the mixture would become saturated in these components, which would precipitate out of the core and rise up to the CMB, possibly providing a buoyancy source for convection in the early core. A species such as Mg would actually precipitate as MgO, which would then react with the core fluid to form (Mg,Fe)O with a high ratio of FeO/FeO + MgO, thus stripping more oxygen from the core than magnesium and amplifying the buoyancy available for this process. The possibility for this kind of mechanism to occur, and the detailed time evolution of precipitation that would result, remains to be investigated.

7.11.8.4 Summary

Chemical disequilibrium between the core and mantle will give rise to reactions along the CMB, although the direction

of the reactions depends not only upon chemical properties but also on the complex evolution of the CMB region. The extent of such reactions depends critically upon the ability for the core and mantle materials to mix over length scales larger than permitted by diffusion alone. The key to understanding the evolution of core–mantle mass exchange through geologic history is improving our understanding of the chemical and physical properties of the core and mantle materials at CMB conditions. Further experimental and theoretical work to better constrain these properties is needed before we can improve our understanding of how the complex features in the CMB region may be produced by these mechanisms.

7.11.9 Concluding Remarks

The dynamical evolution and state of the Earth's CMB region are a very challenging but increasingly tractable geologic problem. This chapter only touches upon some of the major research topics presently under way; however, the field is very broad and continues to evolve significantly on short timescales. Additionally, new frontiers are presently opening, such as establishing connections between the CMB region and the surface environment, as well as the inner core. Like any geologic environment, the present CMB region represents a convolution of a complex thermal and chemical evolution of the Earth since formation, and the grand challenge is to understand how all of the Earth's systems evolved over time to their present state and how they are linked to one another. We look forward to further advances in our understanding in the coming years and anticipate that a major revision of this chapter will be necessary after several years have elapsed.

Acknowledgments

The authors are grateful to Christine Houser for helping with the manuscript preparation; to Edward J. Garnero for contributing one of the figures; to Stephane Labrosse and David Gubbins for their insightful discussions; and to Shijie Zhong, Gerald Schubert, and David Bercovici for their helpful reviews that improved this chapter.

References

Agee CB and Walker D (1988) Static compression and olivine flotation in ultrabasic silicate liquid. *Journal of Geophysical Research B* 93: 3437–3449.

Akber-Knutson S, Steinle-Neumann G, and Asimow PD (2005) Effect of Al on the sharpness of the MgSiOs perovskite to post-perovskite phase transition. *Geophysical Research Letters* 32: L14303. http://dx.doi.org/10.1029/2005GL023192.

Akins JA, Luo S-N, Asimow PD, and Ahrens TJ (2004) Shock-induced melting of $MgSiO_3$ and implications for melts in Earth's lowermost mantle. *Geophysical Research Letters* 31: L14612.

Albarède F (1998) Time-dependent models of UThHe and KAr evolution and the layering of mantle convection. *Chemical Geology* 145: 413–429.

Alexandrakis C and Eaton DW (2010) Precise seismic-wave velocity atop Earth's core: No evidence for outer-core stratification. *Physics of the Earth and Planetary Interiors* 180: 59–65.

Alfé D, Gillan MJ, and Price GD (2002) Composition and temperature of the Earth's core constrained by combining ab initio calculations and seismic data. *Earth and Planetary Science Letters* 195: 91–98.

Ammann MW, Brodholt JP, Wookey J, and Dobson DP (2010) First-principles constraints on diffusion in lower-mantle minerals and a weak *D″* layer *Nature* 465(7297): 462–465.

Andrault D, Munoz M, Bolfan-Casanova N, Guignot N, Perrillat J-P, Aquilanti G, and Pascarelli S (2010) Experimental evidence for perovskite and post-perovskite coexistence throughout the whole *D″* region *Earth and Planetary Science Letters* 293: 90–96.

Andrault D, Bolfan-Casanova N, Lo Nigro G, Bouhifd MA, Garbarino G, and Mezouar M (2011) Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history. *Earth and Planetary Science Letters* 304: 251–259. http://dx.doi.org/10.1016/j.epsl.2011.02.006.

Andrault D, Petitgirard S, Nigro GL, Devidal JL, Veronesi G, Garbarino G, and Mezouar M (2012) Solid–liquid iron partitioning in Earth's deep mantle. *Nature* 487: 354–359.

Antolik M, Gu Y, Ekstrom G, and Dziewonski A (2003) J362D28: A new joint model of compressional and shear velocity in the Earth's mantle. *Geophysical Journal International* 153: 443–466.

Asahara Y, Frost DJ, and Rubie DC (2007) Partitioning of FeO between magnesiowüstite and liquid iron at high pressures and temperatures: Implications for the composition of the Earth's outer core. *Earth and Planetary Science Letters* 257: 435–449. http://dx.doi.org/10.1016/j.epsl.2007.03.006.

Aubert J, Amit H, Hulot G, and Olson P (2008) Thermochemical flows couple the Earth's inner core growth to mantle heterogeneity. *Nature* 454: 758–761. http://dx.doi.org/10.1038/nature07109.

Auzende A-L, Badro J, Ryerson FJ, Weber PK, Fallon SJ, Addad A, Siebert J, and Fiquet G (2008) Element partitioning between magnesium silicate perovskite and ferropericlase: New insights into bulk lower-mantle geochemistry. *Earth and Planetary Science Letters* 269: 164–174. http://dx.doi.org/10.1016/j.epsl.2008.02.001.

Avants M, Lay T, Russell SA, and Garnero EJ (2006) Shear velocity variation within the *D″* region beneath the central Pacific *Journal of Geophysical Research* 111: B05305. http://dx.doi.org/10.1029/2004JB003270.

Badro J, Rueff J-P, Vanko G, Monaco G, Fiquet G, and Guyot F (2004) Electronic transitions in perovskite: Possible nonconvecting layers in the lower mantle. *Science* 305: 383–386.

Beggan C and Whaler K (2008) Core flow modelling assumptions. *Physics of the Earth and Planetary Interiors* 167: 217–222.

Berryman JG (2000) Seismic velocity decrement ratios for regions of partial melt in the lower mantle. *Geophysical Research Letters* 27: 421–424.

Bi Y, Tan H, and Jing F (2002) Electrical conductivity of iron under shock compression up to 200 GPa. *Journal of Physics: Condensed Matter* 14: 10849–10854.

Birch F (1964) Density and composition of the mantle and core. *Journal of Geophysical Research* 69: 4377–4388.

Bloxham J and Jackson A (1991) Fluid flow near the surface of Earth's outer core. *Reviews of Geophysics* 29: 97–120.

Boehler R (2000) High-pressure experiments and the phase diagram of lower mantle and core constituents. *Reviews of Geophysics* 38: 221–245.

Bolton H and Masters G (2001) Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle. *Journal of Geophysical Research* 106: 13527–13540.

Bower DJ, Gurnis M, and Seton M (2013) Lower mantle structure from paleogeographically constrained dynamic Earth models. *Geochemistry, Geophysics, Geosystems* 14(1): 44–63.

Braginsky SI (1963) Structure of the F layer and reasons for convection in the Earth's core. *Doklady Akademii Nauk SSSR* 149: 8–10.

Braginsky SI (2007) Formation of the stratified ocean of the core: A ternary alloy model. *Earth and Planetary Science Letters* 253: 507–512.

Braginsky SI and Roberts PH (1995) Equations governing convection in Earth's core and the geodynamo. *Geophysical and Astrophysical Fluid Dynamics* 79: 1–97.

Brandenburg JP and van Keken PE (2007a) Deep storage of oceanic crust in a vigorously convecting mantle. *Journal of Geophysical Research-Solid Earth* 112(B6).

Brandenburg JP and van Keken PE (2007b) Methods for thermochemical convection in Earth's mantle with force-balanced plates. *Geochemistry, Geophysics, Geosystems* 8.

Brandenburg JP, Hauri EH, van Keken PE, and Ballentine CJ (2008) A multiple-system study of the geochemical evolution of the mantle with force-balanced plates and thermochemical effects. *Earth and Planetary Science Letters* 276(1–2): 1–13.

Brandon AD and Walker RJ (2005) The debate over core–mantle interaction. *Earth and Planetary Science Letters* 232: 211–225.

Breger L and Romanowicz B (1998) Three-dimensional structure at the base of the mantle beneath the central Pacific. *Science* 282(5389): 718–720.

Buffett BA (2002) Estimates of heat flow in the deep mantle based on the power requirements for the geodynamo. *Geophysical Research Letters* 29, GL014649.
Buffett BA (2007) A bound on heat flow below a double crossing of the perovskite–postperovskite phase transition. *Geophysical Research Letters* 34: L17302. http://dx.doi.org/10.1029/2007GL030930.
Buffett BA and Seagle CT (2010) Stratification of the top of the core due to chemical interactions with the mantle. *Journal of Geophysical Research* 115: B04407. http://dx.doi.org/10.1029/2009JB006751.
Buffett BA, Garnero EJ, and Jeanloz R (2000) Sediments at the top of the Earth's core. *Science* 290: 1338–1342.
Bull AL, McNamara AK, and Ritsema J (2009) Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters* 278(3–4): 152–162.
Bull AL, McNamara AK, Becker TW, and Ritsema J (2010) Global scale models of the mantle flow field predicted by synthetic tomography models. *Physics of the Earth and Planetary Interiors* 182(3–4): 129–138.
Bullen KE (1949) Compressibility-pressure hypothesis and the Earth's interior. *Monthly Notices of the Royal Astronomical Society Supplement* 5: 355–368.
Bunge H-P, Richards MA, Lithgow-Bertelloni C, Baumgardner JR, Grand SP, and Romanowicz B (1998) Time scales and heterogeneous structure in geodynamic Earth models. *Science* 280(5360): 91–95.
Burke K and Torsvik TH (2004) Derivation of Large Igneous Provinces of the past 200 million years from long-term heterogeneities in the deep mantle. *Earth and Planetary Science Letters* 227(3–4): 531–538.
Burke K, Steinberger B, Torsvik TH, and Smethurst MA (2008) Plume Generation Zones at the margins of large low shear velocity provinces on the core–mantle boundary. *Earth and Planetary Science Letters* 265: 49–60.
Cadio C, Panet I, Davaille A, Diament M, Metivier L, and de Viron O (2011) Pacific geoid anomalies revisited in light of thermochemical oscillating domes in the lower mantle. *Earth and Planetary Science Letters* 306(1–2): 123–135.
Caldeira K and Rampino MR (1991) The mid-Cretaceous super plume, carbon dioxide, and global warming. *Geophysical Research Letters* 18: 987–990. http://dx.doi.org/10.1029/91GL01237.
Carlson RW (1994) Mechanisms of Earth differentiation: Consequences for the chemical structure of the mantle. *Reviews of Geophysics* 32(4): 337–361.
Catalli K, Shim S-H, and Prakapenka V (2009) Thickness and clapeyron slope of the post-perovskite boundary. *Nature* 462: 782–786. http://dx.doi.org/10.1038/nature08598.
Christensen UR (2006) A deep rooted dynamo for Mercury. *Nature* 444: 1056–1058.
Christensen UR and Hofmann AW (1994) Segregation of subducted oceanic crust in the convecting mantle. *Journal of Geophysical Research B* 99: 19867–19884.
Clark SP (1957) Absorption spectra of some silicates in the visible and near infrared. *American Mineralogist* 43: 732–742.
Clement BM (1991) Geographical distribution of transitional VGPs: Evidence for non-zonal equatorial symmetry during the Matuyama-Brunhes geomagnetic reversal. *Earth and Planetary Science Letters* 104: 48–58.
Cobden L and Thomas C (2013) The origin of *D″* reflections: A systematic study of seismic array data sets. *Geophysical Journal International* 194: 1091–1118. http://dx.doi.org/10.1093/gji/ggt152.
Coltice N and Ricard Y (1999) Geochemical observations and one layer mantle convection. *Earth and Planetary Science Letters* 174: 125–137.
Condie KC (2001) *Mantle Plumes and Their Record in Earth History*. Cambridge University Press.
Conrad CP, Steinberger B, and Torsvik TH (2013) Stability of active mantle upwelling revealed by net characteristics of plate tectonics. *Nature* 498(7455): 479–482.
Cormier VF (2000) *D″* as a transition in the heterogeneity spectrum of the lowermost mantle *Journal of Geophysical Research* 105: 16193–16205.
Costin SO and Buffett BA (2004) Preferred reversal paths caused by a heterogeneous conducting layer at the base of the mantle. *Journal of Geophysical Research* 109: B06101. http://dx.doi.org/10.1029/2003JB002853.
Cottaar S and Romanowicz B (2012) An unusually large ULVZ at the base of the mantle near Hawaii. *Earth and Planetary Science Letters* 355–356: 213–222. http://dx.doi.org/10.1016/j.epsl.2012.09.005.
Courtillot V and Olson P (2007) Mantle plumes link magnetic superchrons to phanerozoic mass depletion events. *Earth and Planetary Science Letters* 260: 495–504. http://dx.doi.org/10.1016/j.epsl.2007.06.003.
Courtillot V, Jaeger JJ, Yang Z, Feraud G, and Hofmann C (1996) The influence of continental flood basalts on mass extinctions: Where do we stand? *GSA Special Papers* 307: 513–525. http://dx.doi.org/10.1130/0-8137-2307-8.513.
Courtillot V, Davaille A, Besse J, and Stock J (2003) Three distinct types of hotspots in the Earth's mantle. *Earth and Planetary Science Letters* 205(3–4): 295–308.
Crameri F, Tackley PJ, Meilick I, Gerya TV, and Kaus BJP (2012) A free plate surface and weak oceanic crust produce single-sided subduction on Earth. *Geophysical Research Letters* 39.
Davaille A (1999) Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* 402: 756–760.
Davaille A, Girard F, and Le Bars M (2002) How to anchor plumes in a convecting mantle? *Earth and Planetary Science Letters* 203: 621–634.
Davaille A, Le Bars M, and Carbonne C (2003) Thermal convection in a heterogeneous mantle. *Comptes Rendus Geoscience* 335: 141–156.
Davaille A, Stutzmann E, Silveira G, Besse J, and Courtillot V (2005) Convective patterns under the Indo-Atlantic box. *Earth and Planetary Science Letters* 239(3–4): 233–252.
Davies DR, Goes S, Davies JH, Schuberth BSA, Bunge H-P, and Ritsema J (2012) Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity. *Earth and Planetary Science Letters* 353–354: 253–269.
Davies GF (1988) Ocean bathymetry and mantle convection: 1. Large-scale flow and hotspots. *Journal of Geophysical Research* 93: 10467–10480.
Davies GF (2008) Episodic layering of the early mantle by the 'basalt barrier' mechanism. *Earth and Planetary Science Letters* 275(3–4): 382–392.
Davies GF and Gurnis M (1986) Interaction of mantle dregs with convection: Lateral heterogeneity at the core mantle boundary. *Geophysical Research Letters* 13: 1517–1520.
de Koker N, Steinle-Neumann G, and Vlcek V (2012) Electrical resistivity and thermal conductivity of liquid Fe alloys at high P and T, and heat flux in Earth's core. *Proceedings of the National Academy of Sciences* 190: 4070–4073.
de Wijs GA, Kresse G, Vocadlo L, Dobson D, Alfe D, Gillan MJ, and Price GD (1998) The viscosity of liquid iron under Earth's core conditions. *Nature* 392: 805–807.
Deguen R, Alboussiere T, and Brito D (2007) On the existence and structure of a mush at the inner core boundary of the Earth. *Physics of the Earth and Planetary Interiors* 164: 36–49. http://dx.doi.org/10.1016/j.pepi.2007.05.003.
Deschamps F and Tackley PJ (2008) Searching for models of thermo-chemical convection that explain probabilistic tomography. I. Principles and influence of rheological parameters. *Physics of the Earth and Planetary Interiors* 171(1–4): 357–373.
Deschamps F and Tackley PJ (2009) Searching for models of thermo-chemical convection that explain probabilistic tomography. II. Influence of physical and compositional parameters. *Physics of the Earth and Planetary Interiors* 176(1–2): 1–18.
Deschamps F and Trampert J (2003) Mantle tomography and its relation to temperature and composition. *Physics of the Earth and Planetary Interiors* 140(4): 277–291.
Deschamps F, Kaminski E, and Tackley PJ (2011) A deep mantle origin for the primitive signature of ocean island basalt. *Nature Geoscience* 4(12): 879–882.
Deschamps F, Cobden L, and Tackley PJ (2012) The primitive nature of large low shear-wave velocity provinces. *Earth and Planetary Science Letters* 349–350: 198208.
Dewaele A, Loubeyre P, Occelli F, Mezouar M, Dorogokupets PI, and Torrent M (2006) Quasihydrostatic equation of state of iron above 2 mbar. *Physical Review Letters* 97: 215504.
Dobson DP and Brodholt JP (2005) Subducted banded iron formations as a source of ultralow-velocity zones at the core–mantle boundary. *Nature* 434: 371–374.
Dobson DP, Miyajima N, Nestola F, Alvaro M, Casati N, Liebske C, Wood IG, and Walker AM (2013) Inherited textures during the perovskite to post-perovskite transition and seismic anisotropy in *D″* *Nature Geoscience* 6: 575–578.
Doornbos DJ and Husebye ES (1972) Array analysis of PKP phases and their precursors. *Physics of the Earth and Planetary Interiors* 5: 387–399.
Driscoll P and Olson P (2011) Superchron cycles driven by variable core heat flow. *Geophysical Research Letters* 38(9).
Dubrovinsky L, Dubrovinskaia N, and Langenhorst F (2004) Reaction of iron and silica at core–mantle boundary conditions. *Physics of the Earth and Planetary Interiors* 146: 243–247.
Duncan RA and Richards MA (1991) Hotspots, mantle plumes, flood basalts, and true polar wander. *Reviews of Geophysics* 29(1): 31–50.
Dziewonski AM and Anderson DL (1981) Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25: 297–356.
Dziewonski AM, Hager BH, and O'Connell RJ (1977) Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research* 82: 239–255.
Dziewonski AM, Lekic V, and Romanowicz B (2010) Mantle anchor structure: An argument for bottom up tectonics. *Earth and Planetary Science Letters* 299(1–2): 69–79.
Eaton DW and Kendall J-M (2006) Improving seismic resolution of outermost core structure by multichannel analysis and deconvolution of broadband SmKS phases. *Physics of the Earth and Planetary Interiors* 155: 104–119.
Engdahl ER, van der Hilst RD, and Berrocal J (1995) Imaging of subducted lithosphere beneath South America. *Geophysical Research Letters* 22(16): 2317–2320.
Farnetani CG (1997) Excess temperature of mantle plumes: The role of chemical stratification across *D″* . *Geophysical Research Letters* 24: 1583–1586.

Farnetani CG and Samuel H (2003) Lagrangian structures and stirring in the Earth's mantle. *Earth and Planetary Science Letters* 206(3–4): 335–348.
Fearn DR and Loper DE (1981) Compositional convection and stratification of Earth's core. *Nature* 289: 393–394.
Fiquet G, Auzende AL, Siebert J, Corgne A, Bureau H, Ozawa H, and Garbarino G (2010) Melting of peridotite to 140 Gigapascals. *Science* 329: 1516. http://dx.doi.org/10.1126/science.1192448.
Fischer RA and Campbell AJ (2010) High pressure melting of wüstite. *American Mineralogist* 95: 1473–1477.
Flores C and Lay T (2005) The trouble seeing double. *Geophysical Research Letters* 32, L24305.
Ford SR, Garnero EJ, and McNamara AK (2006a) A strong lateral shear velocity gradient and anisotropy heterogeneity in the lowermost mantle beneath the southern pacific. *Journal of Geophysical Research-Solid Earth* 111(B3).
Ford SR, Garnero EJ, and McNamara AK (2006b) A strong lateral shear velocity gradient and anisotropy heterogeneity in the lowermost mantle beneath the southern Pacific. *Journal of Geophysical Research* 111: B03306. http://dx.doi.org/10.1029/2004JB003574.
Forte AM and Mitrovica JX (2001) Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data. *Nature* 410: 1049–1056.
Frost DJ, Asahara Y, Rubie DC, Miyajima N, Dubrovinsky LS, Holzapfel C, Ohtani E, Miyahara M, and Sakai T (2010) Partitioning of oxygen between the Earth's mantle and core. *Journal of Geophysical Research* 115: B02202. http://dx.doi.org/10.1029/2009JB006302.
Fukao Y, Obayashi M, Inoue H, and Nenbai M (1992) Subducting slabs stagnant in the mantle transition zone. *Journal of Geophysical Research-Solid Earth* 97(B4): 4809–4822.
Garnero EJ and Helmberger DV (1995) On seismic resolution of lateral heterogeneity in the Earth's outermost core. *Physics of the Earth and Planetary Interiors* 88: 117–130.
Garnero EJ and Helmberger DV (1996) Seismic detection of a thin laterally varying boundary layer at the base of the mantle beneath the central-Pacific. *Geophysical Research Letters* 23: 977–980.
Garnero EJ and McNamara AK (2008) Structure and dynamics of Earth's lower mantle. *Science* 320: 626–628.
Garnero EJ, Helmberger DV, and Grand SP (1993a) Constraining outermost core velocity with SmKS waves. *Geophysical Research Letters* 20: 2463–2466.
Garnero EJ, Helmberger DV, and Grand SP (1993b) Preliminary evidence for a lower mantle shear wave velocity discontinuity beneath the central Pacific. *Physics of the Earth and Planetary Interiors* 79: 335–347.
Glatzmaier GA, Coe RS, Hongre L, and Roberts PH (1999) The role of the Earth's mantle in controlling the frequency of geomagnetic reversals. *Nature* 401: 885–890.
Gomi H, Ohta K, Hirose K, Labrosse S, Caracas R, Verstraete MJ, and Hernlund JW (2013) The high conductivity of iron and thermal evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 224: 88–103.
Goncharov AF, Beck P, Struzhkin VV, Haugen BD, and Jacobsen SD (2007) Thermal conductivity of lower mantle minerals. *Physics of the Earth and Planetary Interiors*, http://dx.doi.org/10.1016/j.pepi.2008.07.033.
Goncharov AF, Haugen BD, Struzhkin VV, Beck P, and Jacobsen SD (2008) Radiative conductivity in the Earth's lower mantle. *Nature* 456: 231–234.
Gonnermann HM and Mukhopadhyay S (2007) Non-equilibrium degassing and a primordial source for helium in ocean-island volcanism. *Nature* 449(7165): 1037–1040.
Gonnermann HM and Mukhopadhyay S (2009) Preserving noble gases in a convecting mantle. *Nature* 459(7246): 560–563.
Grand SP (1994) Mantle shear structure beneath the Americas and surrounding oceans. *Journal of Geophysical Research* 99: 11591–11621.
Grand SP (2002) Mantle shear-wave tomography and the fate of subducted slabs. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 360(1800): 2475–2491.
Grand SP, van der Hilst RD, and Widiyantoro S (1997) Global seismic tomography: A snapshot of convection in the Earth. *GSA Today* 7: 1–7.
Grocholski B, Catalli K, Shim S-H, and Prakapenka V (2012) Mineralogical effects on the detectability of the postperovskite boundary. *Proceedings of the National Academy of Sciences*, http://dx.doi.org/10.1073/pnas.1109204109.
Gu YJ, Dziewonski AM, Su WJ, and Ekstrom G (2001) Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities. *Journal of Geophysical Research, Solid Earth* 106(B6): 11169–11199.
Gubbins D (2007a) Geomagnetic constraints on stratification at the top of the Earth's core. *Earth, Planets and Space* 59: 661–664.
Gubbins D and Davies CJ (2013) The stratified layer at the core–mantle boundary caused by barodiffusion of oxygen, sulphur and silicon. *Physics of the Earth and Planetary Interiors* 215: 21–28.
Gubbins D, Alfe D, Masters G, Price GD, and Gillan MJ (2003) Can the Earth's dynamo run on heat alone? *Geophysical Journal International* 155: 609–622.
Gubbins D, Willis AP, and Sreenivasan B (2007) Correlation of Earth's magnetic field with lower mantle thermal and seismic structure. *Physics of the Earth and Planetary Interiors* 162: 256–260.
Gunnarsson O, Calandra M, and Han JE (2003) Colloquium: Saturation of electrical resistivity. *Reviews of Modern Physics* 75: 1085–1099.
Gurnis M (1986) The effects of chemical density differences on convective mixing in the Earth's mantle. *Journal of Geophysical Research B* 91: 1407–1419.
Gurnis M and Hager BH (1988) Controls of the structure of subducted slabs. *Nature* 335(6188): 317–321.
Hager BH (1984) Subducted slabs and the geoid: Constraints on mantle rheology and flow. *Journal of Geophysical Research* 89: 6003–6015. http://dx.doi.org/10.1029/JB089iB07p06003.
Hager BH and Richards MA (1989) Long-wavelength variations in Earth's geoid: Physical models and dynamical implications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 328(1599): 309–327.
Hales AL and Roberts JL (1971) The velocities in the outer core. *Bulletin of the Seismological Society of America* 61: 1051–1059.
Hansen U and Yuen DA (1988) Numerical simulations of thermal-chemical instabilities at the core mantle boundary. *Nature* 334: 237–240.
He Y and Wen LX (2012) Geographic boundary of the Pacific Anomaly and its geometry and transitional structure in the north. *Journal of Geophysical Research* 117: B09308. http://dx.doi.org/10.1029/2012JB009436.
He YM and Wen LX (2009) Structural features and shear-velocity structure of the "pacific anomaly" *Journal of Geophysical Research-Solid Earth* 114.
He YM, Wen LX, and Zheng TY (2006) Geographic boundary and shear wave velocity structure of the "pacific anomaly" near the core–mantle boundary beneath western pacific. *Earth and Planetary Science Letters* 244(1–2): 302–314.
Helffrich G (2012) How light element addition can lower core liquid wave speeds. *Geophysical Journal International* 188: 1065–1070. http://dx.doi.org/10.1111/j.1365-246X.2011.05295.x.
Helffrich G and Kaneshima S (2004) Seismological constraints on core composition from Fe–O–S liquid immiscibility. *Science* 306: 2239–2242.
Helffrich G and Kaneshima S (2010) Outer-core compositional stratification from observed core wave speed profiles. *Nature* 468: 807–810. http://dx.doi.org/10.1038/nature09636.
Helffrich G and Kaneshima S (2013) Causes and consequences of outer core stratification. *Physics of the Earth and Planetary Interiors* 223: 2–7. http://dx.doi.org/10.1016/j.pepi.2013.07.005.
Helmberger D, Ni S, Wen LX, and Ritsema J (2000) Seismic evidence for ultralow-velocity zones beneath Africa and eastern Atlantic. *Journal of Geophysical Research* 105: 23865–23878.
Helmberger DV, Wen L, and Ding X (1998) Seismic evidence that the source of the Iceland hotspot lies at the core–mantle boundary. *Nature* 396: 251–255.
Hemley RJ and Kubicki JD (1991) Deep mantle melting. *Nature* 349: 283–284.
Hernlund J (2013) Mantle fabric unravelled? *Nature Geoscience* 6: 516–518. http://dx.doi.org/10.1038/ngeo1868.
Hernlund JW (2010) On the interaction of the geotherm with a post-perovskite phase transition in the deep mantle. *Physics of the Earth and Planetary Interiors* 180: 222–234.
Hernlund JW and Houser C (2008) On the distribution of seismic velocities in Earth's deep mantle. *Earth and Planetary Science Letters* 265: 423–437. http://dx.doi.org/10.1016/j.epsl.2007.10.042.
Hernlund JW and Jellinek M (2010) Dynamics and structure of a stirred partially molten ultralow-velocity zone. *Earth and Planetary Science Letters* 296: 1–8. http://dx.doi.org/10.1016/j.epsl.2010.04.027.
Hernlund JW and Labrosse S (2007) Geophysically consistent values of the perovskite to post-perovskite transition Clapeyron slope. *Geophysical Research Letters* 34: L05309. http://dx.doi.org/10.1029/2006GL028961.
Hernlund JW and Tackley PJ (2007) Some dynamical consequences of partial melting at the base of Earth's mantle. *Physics of the Earth and Planetary Interiors* 162: 149–163.
Hernlund JW, Thomas C, and Tackley PJ (2005) A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* 434: 882–886. http://dx.doi.org/10.1038/nature03472.
Herzberg C, Condie K, and Korenaga J (2010) Thermal history of the Earth and its petrological expression. *Earth and Planetary Science Letters* 292: 79–88.

Hier-Majumder S (2004) Influence of contiguity on seismic velocities of partially molten aggregates. *Journal of Geophysical Research* 113: B12205, http://dx.doi.org/10.1029/2008JB005662.
Hirose K, Fei YW, Ma YZ, and Mao HK (1999) The fate of subducted basaltic crust in the Earth's lower mantle. *Nature* 397(6714): 53–56.
Hirose K, Takafuji N, Sata N, and Ohishi Y (2005) Phase transition and density of subducted MORB crust in the lower mantle. *Earth and Planetary Science Letters* 237(1–2): 239–251.
Hirose K, Sinmyo R, Sata N, and Ohishi Y (2006) Determination of post-perovskite phase transition in $MgSiO_3$ using Au and MgO pressure standards. *Geophysical Research Letters* 33: L01310.
Hirose K, Labrosse S, and Hernlund J (2013) Composition and state of the core. *Annual Review of Earth and Planetary Sciences* 41: 657–691. http://dx.doi.org/10.1146/annurev-earth-050212-124007.
Hofmann AW (1997) Mantle geochemistry: The message from oceanic volcanism. *Nature* 385: 219–229.
Hofmann AW and White WM (1982) Mantle plumes from ancient oceanic crust. *Earth and Planetary Science Letters* 57: 421–436.
Hofmeister AM (1999) Mantle values of thermal conductivity and the geotherm from phonon lifetimes. *Science* 283: 1699–1706.
Houser C, Masters G, Shearer P, and Laske G (2008) Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms. *Geophysical Journal International* 174: 195–212. http://dx.doi.org/10.1111/j.1365-246X.2008.03763.x.
Humayun M, Qin L, and Norman MD (2004) Geochemical evidence for excess iron in the mantle beneath hawaii. *Science* 306: 91. http://dx.doi.org/10.1126/science.1101050.
Hunt SA, Weidner DJ, Li L, Wang L, Walte NP, Brodholt JP, and Dobson DP (2009) Weakening of calcium iridate during its transformation from perovskite to post-perovskite. *Nature Geoscience* 2: 794–797. http://dx.doi.org/10.1038/ngeo663.
Hunt SA, Davies DR, Walker AM, McCormack RJ, Wills AS, Dobson DP, and Li L (2012) On the increase in thermal diffusivity caused by the perovskite to post-perovskite phase transition and its implications for mantle dynamics. *Earth and Planetary Science Letters* 319320: 96–103.
Hutko AR, Lay T, Revenaugh J, and Garnero EJ (2008) Anticorrelated seismic velocity anomalies from post-perovskite in the lowermost mantle. *Science* 320: 1070–1074. http://dx.doi.org/10.1126/science.1155822.
Ishii M and Tromp J (1999) Normal-mode and free-air gravity constraints on lateral variations in velocity and density of the Earth's mantle. *Science* 285: 1231–1236.
Ishii M and Tromp J (2001) Even-degree lateral variations in the Earth's mantle constrained by free oscillations and the free-air gravity anomaly. *Geophysical Journal International* 145(1): 77–96.
Ishii M and Tromp J (2004a) Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes. *Physics of the Earth and Planetary Interiors* 146: 113–124.
Ishii M and Tromp J (2004b) Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive modes. *Physics of the Earth and Planetary Interiors* 146: 113–124.
Jaupart C, Labrosse S, and Mareschal J-C (2007) Temperatures, heat and energy in the mantle of the Earth. In: Bercovici D and Schubert G (eds.) *Treatise on Geophysics*, pp. 253–303. Amsterdam: Elsevier.
Jellinek AM and Manga M (2002) The influence of a chemical boundary layer on the fixity and lifetime of mantle plumes. *Nature* 418: 760–763.
Jellinek AM and Manga M (2004) Links between long-lived hotspots, mantle plumes, *D″*, and plate tectonics *Reviews of Geophysics* 42: RG3002. http://dx.doi.org/10.1029/2003RG000144.
Jellinek AM, Gonnermann HM, and Richards MA (2003) Plume capture by divergent plate motions: Implications for the distribution of hotspots, geochemistry of mid-ocean ridge basalts, and estimates of the heat flux at the core–mantle boundary. *Earth and Planetary Science Letters* 205: 361–378.
Jensen KJ, Thorne MS, and Rost S (2013) SPdKS analysis of ultralow-velocity zones beneath the western Pacific. *Geophysical Research Letters* 40: 1–5. http://dx.doi.org/10.1002/grl.50877.
Kanda RVS and Stevenson DJ (2006) Suction mechanism for iron entrainment into the lower mantle. *Geophysical Research Letters* 33: L02310. http://dx.doi.org/10.1029/2005GL025009.
Kawai K and Tsuchiya T (2009) Temperature profile in the lowermost mantle from seismological and mineral physics joint modeling. *Proceedings of the National Academy of Sciences of the United States of America* 106: 22119–22123. http://dx.doi.org/10.1073/pnas.0905920106.
Kawai K, Geller RJ, and Fuji N (2004a) *D″* beneath the Arctic from inversion of shear waveforms *Geophysical Research Letters* 34: L21305. http://dx.doi.org/10.1029/2007GL031517.
Kawai K, Takeuchi N, Geller RJ, and Fuji N (2007b) Possible evidence for a double crossing phase transition in *D″* beneath Central America from inversion of seismic waveforms *Geophysical Research Letters* 34: L09314. http://dx.doi.org/10.1029/2007GL029642.
Keeler RN and Mitchell AC (1969) Electrical conductivity, demagnetization, and the high-pressure phase transition in shock-compressed iron. *Solid State Communications* 7: 271–274.
Keller G (2005) Biotic effects of late Maastrichtian mantle plume volcanism: Implications for impacts and mass extinctions. *Lithos* 79: 317–341.
Kellogg LH and King SD (1993) Effect of mantle plumes on the growth of *D″* by reaction between the core and mantle *Geophysical Research Letters* 20: 379–382.
Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283(5409): 1881–1884.
Kennett BLN and Engdahl ER (1991) Traveltimes for global earthquake location and phase identification. *Geophysical Journal International* 105: 429–465.
Kennett BLN, Engdahl ER, and Buland R (1995) Constraints on seismic velocities in the Earth from travel times. *Geophysical Journal International* 122: 108–124.
King SD (1997) Geoid and topographic swells over temperature-dependent thermal plumes in spherical-axisymmetric geometry. *Geophysical Research Letters* 24(23): 3093–3096.
Knittle E and Jeanloz R (1991) The Earth's core–mantle boundary: Results of experiments at high pressures and temperatures. *Science* 251: 1438–1443.
Komabayashi T (2014) Thermodynamics of melting relations in the system Fe–FeO: Implications for the oxygen content in the Earth's outer core. *Journal of Geophysical Research* 119: 4164–4177. http://dx.doi.org/10.1002/2014JB010980.
Kuo BY, Garnero EJ, and Lay T (2000) Tomographic inversion of S-SKS times for shear velocity heterogeneity in *D″*: Degree 12 and hybrid models *Journal of Geophysical Research-Solid Earth* 105(B12): 28139–28157.
Labrosse S (2002) Hotspots, mantle plumes and core heat loss. *Earth and Planetary Science Letters* 199: 147–156.
Labrosse S (2003) Thermal and magnetic evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 140: 127–143.
Labrosse S, Poirier J-P, and Le Mouël J-L (1997) On cooling of the Earth's core. *Physics of the Earth and Planetary Interiors* 99: 1–17.
Labrosse S, Hernlund JW, and Coltice N (2007) A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* 450: 866–869.
Laj C, Mazaud A, Weeks R, Fuller M, and Herrero-Bervera E (1991) Geomagnetic reversal paths. *Nature* 351: 447.
Larson R (1991) Latest pulse of Earth: Evidence for a mid-Cretaceous superplume. *Geology* 19: 547–550.
Lassak TM, McNamara AK, and Zhong S (2007) Influence of thermochemical piles on topography at Earth's core–mantle boundary. *Earth and Planetary Science Letters* 261(3–4): 443–455.
Lassak TM, McNamara AK, Garnero EJ, and Zhong SJ (2010) Core–mantle boundary topography as a possible constraint on lower mantle chemistry and dynamics. *Earth and Planetary Science Letters* 289(1–2): 232–241.
Lay T (1987) Structure of the Earth: Mantle and core. *Reviews of Geophysics* 25(6): 1161–1167.
Lay T (2007) Deep earth structure-lower mantle and *D″* In: Schubert G (ed.) *Treatise on Geophysics*, pp. 619–654. Elsevier.
Lay T and Garnero EJ (2011) Deep mantle seismic modeling and imaging. *Annual Review of Earth and Planetary Sciences* 39: 91–123.
Lay T and Helmberger DV (1983) A shear velocity discontinuity in the lower mantle. *Geophysical Research Letters* 10: 63–66.
Lay T and Young CJ (1990) The stably-stratified outermost core revisited. *Geophysical Research Letters* 17: 2001–2004.
Lay T, Williams Q, Garnero EJ, Kellogg L, and Wysession ME (1998) Seismic wave anisotropy in the *D″* region and its implications In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region*. *American Geophysical Union Monograph*, pp. 299–318.
Lay T, Garnero EJ, and Williams Q (2004) Partial melting in a thermo-chemical boundary layer at the base of the mantle. *Physics of the Earth and Planetary Interiors* 146: 441–467.
Lay T, Hernlund J, Garnero EJ, and Thorne MS (2006) A post-perovskite lens and *D″* heat flux beneath the central Pacific *Science* 314: 1272–1276. http://dx.doi.org/10.1126/science.1133280.
Le Bars M and Davaille A (2002) Stability of thermal convection in two superimposed miscible viscous fluids. *Journal of Fluid Mechanics* 471: 339–363.
Le Bars M and Davaille A (2004a) Whole layer convection in a heterogeneous planetary mantle. *Journal of Geophysical Research-Solid Earth* 109(B3).
Le Bars M and Davaille A (2004b) Large interface deformation in two-layer thermal convection of miscible viscous fluids. *Journal of Fluid Mechanics* 499: 75–110.

Lee C-T, Luffi P, Hoink T, Li J, Dasgupta R, and Hernlund JW (2010) Upside-down differentiation and generation of a primordial lower mantle. *Nature* 463: 930–935.
Lekic V, Cottaar S, Dziewonski AM, and Romanowicz B (2012) Cluster analysis of global lower mantle tomography: A new class of structure and implications for chemical heterogeneity. *Earth and Planetary Science Letters* 357–358: 68–77.
Li C, van der Hilst RD, Engdahl ER, and Burdick S (2008) A new global model for P wave speed variations in Earth's mantle. *Geochemistry, Geophysics, Geosystems* 9.
Li J and Agee CB (1996) Geochemistry of mantle–core differentiation at high pressure. *Nature* 381: 686–689. http://dx.doi.org/10.1038/381686a0.
Li M and McNamara AK (2013) The difficulty for subducted oceanic crust to accumulate at the Earth's core–mantle boundary. *Journal of Geophysical Research, Solid Earth* 1–10.
Li M, McNamara AK, and Garnero EJ (2014) Chemical complexity of hotspots caused by cycling oceanic crust through mantle reservoirs. *Nature Geoscience.* http://dx.doi.org/10.1038/ngeo2120.
Li XD and Romanowicz B (1996) Global mantle shear velocity model developed using nonlinear asymptotic coupling theory. *Journal of Geophysical Research-Solid Earth* 101(B10): 22245–22272.
Lithgow-Bertelloni C and Richards MA (1998) The dynamics of Cenozoic and Mesozoic plate motions. *Reviews of Geophysics* 36: 27–78.
Liu L and Bassett WA (1986) *Elements, Oxides, Silicates: High-Pressure Phases with Implications for the Earth's Interior*. Oxford University Press.
Loper DE (1978) Some thermal consequences of a gravitationally powered dynamo. *Journal of Geophysical Research* 83: 5961–5970.
Lord OT, Walter MJ, Dobson DP, Armstrong L, Clark SM, and Kleppe A (2010) The FeSi phase diagram to 150 GPa. *Journal of Geophysical Research* 115: B06208. http://dx.doi.org/10.1029/2009JB006528.
Luo S-N, Ni SD, and Helmberger DV (2001) Evidence for a sharp lateral variation of velocity at the core–mantle boundary from multipathed PKPab. *Earth and Planetary Science Letters* 189: 155–164.
Manga M and Jeanloz R (1996) Implications of a metal-bearing chemical boundary layer in *D″* for mantle dynamics *Geophysical Research Letters* 23: 3091–3094.
Manga M and Jeanloz R (1997) Thermal conductivity of corundum and periclase, implications for the lower mantle. *Journal of Geophysical Research* 102: 2999–3008.
Mann U, Frost DJ, and Rubie DC (2009) Evidence for high-pressure core–mantle differentiation from the metal-silicate partitioning of lithophile and weakly-siderophile elements. *Geochimica et Cosmochimica Acta* 73: 7360–7386.
Manners U (2008) *Investigating the Structure of the Core–Mantle Boundary Using S and P Diffracted Waves*. San Diego: University of California.
Manthilake MAGM, de Koker N, and Frost DJ (2011) Thermal conductivity of $CaGeO_3$ perovskite at high pressure. *Geophysical Research Letters* 38: L08301. http://dx.doi.org/10.1029/2011GL046882.
Mao HK and Bell PM (1972) Electrical conductivity and the red shift of absorption in olivine and spinel at high pressure. *Science* 176: 403–406.
Mao WL, Shen G, Prakapenka VB, Meng Y, Cambell AJ, Heinz D, Shu J, Hemley RJ, and Mao HK (2004) Ferromagnesian postperovskite silicates in the *D″* layer of the Earth *Proceedings of the National Academy of Sciences of the United States of America* 101: 15867–15869.
Mao WL, Mao H-K, Sturhahn W, Zhao J, Prakapenka VB, Shu J, Fei Y, and Hemley RJ (2006) Iron-rich post-perovskite and the origin of ultralow-velocity zones. *Science* 312: 564–565. http://dx.doi.org/10.1126/science.1123442.
Masters G, Johnson S, Laske G, and Bolton H (1996) A shear-velocity model of the mantle. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 354: 1385–1411.
Masters G, Laske G, Bolton H, and Dziewonski A (2000) The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure. In: Karato S-I, Forte A, Liebermann R, Masters G, and Stixrude L (eds.) *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to Global Scale. American Geophysical Union Monograph*, pp. 63–87. American Geophysical Union.
Matassov G (1977) *The Electrical Conductivity of Iron-Silicon Alloys at High Pressures and the Earth's Core*. Rep UCRL-52322, Lawrence Livermore National Lab.
Matyska C, Moser J, and Yuen DA (1994) The potential influence of radiative heat transfer on the formation of megaplumes in the lower mantle. *Earth and Planetary Science Letters* 125: 255–266.
McCammon C, Kantor I, Narygina O, Rouquette J, Ponkratz U, Sergueev I, Mezouar M, Prakapenka V, and Dubrovinsky L (2008) Stable intermediate-spin ferrous iron in lower-mantle perovskite. *Nature Geoscience* 1: 684–687. http://dx.doi.org/10.1038/ngeo309.
McNamara AK and van Keken PE (2000) Cooling of the Earth: A parameterized convection study of whole versus layered models. *Geochemistry, Geophysics, Geosystems* 1. http://dx.doi.org/10.1029/2000GC000045.
McNamara AK and Zhong S (2004) Thermochemical structures within a spherical mantle: Superplumes or piles? *Journal of Geophysical Research B* 109: B07402.
McNamara AK and Zhong S (2005) Thermochemical structures under Africa and the Pacific Ocean. *Nature* 437: 1136–1139.
McNamara AK, Karato S, and van Keken PE (2001) Localization of dislocation creep in the lower mantle: Implications for the origin of seismic anisotropy. *Earth and Planetary Science Letters* 191: 85–99.
McNamara AK, van Keken PE, and Karato S (2003) Lattice-preferred orientation near the core-mantle boundary; A likely mechanism to produce seismic anisotropy. *Journal of Geophysical Research* 108: 2230. http://dx.doi.org/10.1029/2002JB001970.
McNamara AK, Garnero EJ, and Rost S (2010) Tracking deep mantle reservoirs with ultra-low velocity zones. *Earth and Planetary Science Letters* 299(1–2): 1–9.
McNutt MK (1998) Superswells. *Superswells. Reviews of Geophysics* 36(2): 211–244.
Meade C, Mao HK, and Hu J (1995) High-temperature phase transition and dissociation of (Mg, Fe)SiO_3 perovskite at lower mantle pressures. *Science* 268: 1743–1745. http://dx.doi.org/10.1126/science.268.5248.1743.
Megnin C and Romanowicz B (2000) The three-dimensional shear velocity structure of the mantle from the inversion of body, surface, and higher-mode waveforms. *Geophysical Journal International* 143: 709–728.
Miles JW (1961) On the stability of heterogeneous shear flows. *Journal of Fluid Mechanics* 10: 496508.
Mitchell BJ and Helmberger DV (1973) Shear velocities at the base of the mantle from observations of S and ScS. *Journal of Geophysical Research* 78: 6009–6020.
Mittelstaedt E and Tackley PJ (2005) Plume heat flow is much less than CMB heat flow. *Earth and Planetary Science Letters* 241: 202–210.
Miyagi L, Kanitpanyacharoen W, Stackhouse S, Militzer B, and Wenk HR (2011) The enigma of post-perovskite anisotropy: Deformation versus transformation textures. *Physics and Chemistry of Minerals* 38: 665–678. http://dx.doi.org/10.1029/2002JB001970.
Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl R, and Hung S-H (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303: 338–343. http://dx.doi.org/10.1126/science.1092485.
Morelli A and Dziewonski AM (1993) Body wave travel-times and a spherically symmetric P- and S-wave velocity model. *Geophysical Journal International* 112: 178–194.
Morgan WJ (1971) Convection plumes in the lower mantle. *Nature* 230: 42–43.
Mosenfelder JL, Asimow PD, Frost DJ, Rubie DC, and Ahrens TJ (2009) The $MgSiO_3$ system at high pressure: Thermodynamic properties of perovskite, postperovskite, and melt from global inversion of shock and static compression data. *Journal of Geophysical Research* 114: B01203. http://dx.doi.org/10.1029/2008JB005900.
Mukhopadhyay S (2012) Early differentiation and volatile accretion recorded in deep-mantle neon and xenon. *Nature* 486(7401): 101–104.
Murakami M, Hirose K, Sata N, Ohishi Y, and Kawamura K (2004) Phase transition of $MgSiO_3$ perovskite in the deep lower mantle. *Science* 304: 855–858.
Murakami M, Sinogeikin SV, Bass JD, Sata N, Ohishi Y, and Hirose K (2007) Sound velocity of $MgSiO_3$ post-perovskite phase: A constraint on the *D″* discontinuity *Earth and Planetary Science Letters* 259: 18–23.
Nakagawa T (2011) Effect of a stably stratified layer near the outer boundary in numerical simulations of a magnetohydrodynamic dynamo in a rotating spherical shell and its implications for Earth's core. *Physics of the Earth and Planetary Interiors* 187: 342–352. http://dx.doi.org/10.1016/j.pepi.2011.06.001.
Nakagawa T and Tackley PJ (2004a) Effects of thermo-chemical mantle convection on the thermal evolution of the Earth's core. *Earth and Planetary Science Letters* 220(1–2): 107–119.
Nakagawa T and Tackley PJ (2004b) Thermo-chemical structure in the mantle arising from a three-component convective system and implications for geochemistry. *Physics of the Earth and Planetary Interiors* 146: 125–138.
Nakagawa T and Tackley PJ (2005) The interaction between the post-perovskite phase change and a thermo-chemical boundary layer near the core–mantle boundary. *Earth and Planetary Science Letters* 238: 204–216.
Nakagawa T and Tackley PJ (2006) Three-dimensional structures and dynamics in the deep mantle: Effects of post-perovskite phase change and deep mantle layering. *Geophysical Research Letters* 33: L12S11. http://dx.doi.org/10.1029/2006GL025719.
Nakagawa T and Tackley PJ (2008) Lateral variations in CMB heat flux and deep mantle seismic velocity caused by a thermal-chemical-phase boundary layer in 3D spherical convection. *Earth and Planetary Science Letters* 271: 348–358.
Nakagawa T and Tackley PJ (2011) Effects of low-viscosity post-perovskite on thermo-chemical mantle convection in a 3-D spherical shell. *Geophysical Research Letters* 38.

Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2010) The influence of MORB and harzburgite composition on thermo-chemical mantle convection in a 3-D spherical shell with self-consistently calculated mineral physics. *Earth and Planetary Science Letters* 296(3–4): 403–412.

Ni S and Helmberger DV (2001) Probing an ultra-low velocity zone at the core mantle boundary with P and S waves. *Geophysical Research Letters* 28: 2345–2348.

Ni S, Tan E, Gurnis M, and Helmberger D (2002) Sharp sides to the African superplume. *Science* 296: 1850–1852.

Ni SD and Helmberger DV (2003a) Ridge-like lower mantle structure beneath South Africa. *Journal of Geophysical Research-Solid Earth* 108(B2).

Ni SD and Helmberger DV (2003b) Seismological constraints on the South African superplume; could be the oldest distinct structure on Earth. *Earth and Planetary Science Letters* 206(1–2): 119–131.

Ni SD and Helmberger DV (2003c) Further constraints on the African superplume structure. *Physics of the Earth and Planetary Interiors* 140(1–3): 243–251.

Ni SD, Helmberger DV, and Tromp J (2005) Three-dimensional structure of the African superplume from waveform modelling. *Geophysical Journal International* 161(2): 283–294.

Nomura R, Ozawa H, Tateno S, Hirose K, Hernlund J, Muto S, Ishii H, and Hiraoka N (2011) Spin crossover and iron-rich silicate melt in the Earth's deep mantle. *Nature* 473(7346): 199–202. http://dx.doi.org/10.1038/nature09940.

Nowacki A, Walker A, Wookey J, and Kendall J-M (2013) Evaluating post-perovskite as a cause of *D″* anisotropy in regions of palaeosubduction. *Geophysical Journal International* 192: 1085–1090. http://dx.doi.org/10.1093/gji/ggs068.

Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of $MgSiO_3$ in Earth's *D″* layer. *Nature* 430: 445–448.

Oganov AR, Martoak R, Laio A, Raiteri P, and Parrinello M (2005) Anisotropy of Earth's *D″* layer and stacking faults in the $MgSiO_3$ post-perovskite phase. *Nature* 438: 1142–1144. http://dx.doi.org/10.1038/nature04439.

Ohta K, Hirose K, and Hernlund J (2012) Lattice thermal conductivity of $MgSiO_3$ perovskite and post-perovskite at the core–mantle boundary. *Earth and Planetary Science Letters* 349350: 109–115.

Ohtani E and Maeda M (2001) Density of basaltic melt at high pressure and stability of the melt at the base of the lower mantle. *Earth and Planetary Science Letters* 193: 69–75.

Oldham D and Davies JH (2004) Numerical investigation of layered convection in a three-dimensional shell with application to planetary mantles. *Geochemistry, Geophysics, Geosystems* 5.

Olson P and Kincaid C (1991) Experiments on the interaction of thermal-convection and compositional layering at the base of the mantle. *Journal of Geophysical Research B* 96: 4347–4354.

Olson P, Schubert G, and Anderson C (1987) Plume formation in the *D″*-layer and the roughness of the core–mantle boundary. *Nature* 327: 409–413.

Olson P, Schubert G, and Anderson C (1993) Structure of axisymmetric mantle plumes. *Journal of Geophysical Research* 98(B4): 6829–6844.

Ono S (2008) Experimental constraints on the temperature profile in the lower mantle. *Physics of the Earth and Planetary Interiors* 170: 267–273. http://dx.doi.org/10.1016/j.pepi.2008.06.033.

Ono S and Oganov AR (2005) In situ observations of phase transition between perovskite and $CaIrO_3$-type phase in $MgSiO_3$ and pyrolitic mantle composition. *Earth and Planetary Science Letters* 236: 914–932.

Otsuka K and Karato S (2012) Deep penetration of molten iron into the mantle caused by a morphological instability. *Nature* 492: 243–246. http://dx.doi.org/10.1038/nature11663.

Ozawa H, Hirose K, Mitome M, Bando Y, Sata N, and Ohishi Y (2008) Chemical equilibrium between ferropericlase and molten iron to 134 GPa and implications for iron content at the bottom of the mantle. *Geophysical Research Letters* 35: L05308. http://dx.doi.org/10.1029/2007GL032648.

Ozawa H, Hirose K, Mitome M, Bando Y, Sata N, and Ohishi Y (2009) Experimental study of reaction between perovskite and molten iron to 146 GPa and implications for chemical equilibrium at the core–mantle boundary. *Physics and Chemistry of Minerals* 36: 355–363.

Panning M and Romanowicz B (2006) A three-dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophysical Journal International* 167: 361–379. http://dx.doi.org/10.1111/j.1365-246X.2006.03100.x.

Persh SE, Vidale JE, and Earle PS (2001) Absence of short-period ULVZ precursors to PcP and ScP from two regions of the CMB. *Geophysical Research Letters* 28: 859–862.

Petford N, Yuen D, Rushmer T, Brodholt J, and Stackhouse S (2005) Shear-induced material transfer across the core–mantle boundary aided by the post-perovskite phase transition. *Earth, Planets and Space* 57: 459–464.

Poirier JP, Malavergne V, and Le Mouël JL (1998) Is there a thin electrically conducting layer at the base of the mantle? In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region. American Geophysical Union Monograph*, pp. 131–137.

Pozzo M, Davies C, Gubbins D, and Alfè D (2012) Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* 485: 355–358.

Pozzo M, Davies C, Gubbins D, and Alfe D (2013) Transport properties for liquid silicon-oxygen-iron mixtures at Earth's core conditions. *Physical Review B* 87: 014110.

Reasoner C and Revenaugh J (2000) ScP constraints on ultralow-velocity zone density and gradient thickness beneath the Pacific. *Journal of Geophysical Research B* 105: 173–182.

Resovsky JS and Ritzwoller MH (1999) Regularization uncertainty in density models estimated from normal mode data. *Geophysical Research Letters* 26: 2319–2322.

Ricard Y, Richards M, LithgowBertelloni C, and Le Stunff Y (1993) A geodynamic model of mantle density heterogeneity. *Journal of Geophysical Research* 98: 21895–21909.

Richards MA and Engebretson DC (1992) Large-scale mantle convection and history of subduction. *Nature* 30: 437–440.

Ricolleau A, Fei Y, Corgne A, Siebert J, and Badro J (2011) Oxygen and silicon contents of earth's core from high pressure metal-silicate partitioning experiments. *Earth and Planetary Science Letters* 310: 409–421.

Rigden SM, Ahrens TJ, and Stolper EM (1984) Densities of liquid silicates at high pressures. *Science* 226: 1071–1073.

Ringwood AE and Irifune T (1988) Nature of the 650-km seismic discontinuity: Implications for mantle dynamics and differentiation. *Nature* 331(6152): 131–136.

Ritsema J and van Heijst H (2002) Constraints on the correlation of p- and s-wave velocity heterogeneity in the mantle from p, pp, ppp, and *pkpab* traveltimes. *Geophysical Journal International* 149: 482–489.

Ritsema J, Garnero EJ, and Lay T (1997) A strongly negative shear velocity gradient and lateral variability in the lowermost mantle beneath the Pacific. *Journal of Geophysical Research* 102: 20395–20411.

Ritsema J, van Heijst HJ, and Woodhouse JH (1999) Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* 286(5446): 1925–1928.

Ritsema J, van Heijst HJ, and Woodhouse JH (2004) Global transition zone tomography. *Journal of Geophysical Research-Solid Earth* 109(B2).

Ritsema J, McNamara AK, and Bull AL (2007) Tomographic filtering of geodynamic models: Implications for model interpretation and large-scale mantle structure. *Journal of Geophysical Research-Solid Earth* 112(B1).

Ritsema J, van Heijst HJ, Deuss A, and Woodhouse JH (2011) S40RTS: A degree-40 shear velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltimes, and normal-mode splitting function measurements. *Geophysical Journal International* 184: 1223–1236.

Romanowicz B (2003) Global mantle tomography: Progress status in the last 10 years. *Annual Review of Earth and Planetary Sciences* 31: 303–328.

Romanowicz B and Gung Y (2002) Superplumes from the core–mantle boundary to the lithosphere: Implications for heat flux. *Science* 296: 513–516.

Ross AR, Thybo H, and Solodilov LN (2004) Reflection seismic profiles of the core–mantle boundary. *Journal of Geophysical Research B* 109, B08303.

Rost S and Revenaugh J (2001) Seismic detection of rigid zones at the top of the core. *Science* 294: 1911–1914.

Rost S, Garnero EJ, Williams Q, and Manga M (2005) Seismic constraints on a possible plume root at the core–mantle boundary. *Nature* 435: 666–669.

Rost S, Garnero EJ, and Williams Q (2006) Fine-scale ultralow-velocity zone structure from high-frequency seismic array data. *Journal of Geophysical Research* 111: B09310. http://dx.doi.org/10.1029/2005JB004088.

Rost S, Garnero EJ, Thorne MS, and Hutko AR (2010) On the absence of an ultralow-velocity zone in the North Pacific. *Journal of Geophysical Research* 115: B04312. http://dx.doi.org/10.1029/2009JB006420.

Rubie DC, Gessmann CK, and Frost DJ (2004) Partitioning of oxygen during core formation on the Earth and Mars. *Nature* 429: 58–61.

Sakai T, Kondo T, Ohtani E, Terasaki H, Endo N, Kuba T, Suzuki T, and Kikegawa T (2006) Interaction between iron and post-perovskite at core–mantle boundary and core signature in plume source region. *Geophysical Research Letters* 33: L15317. http://dx.doi.org/10.1029/2006GL026868.

Samuel H and Bercovici D (2006) Oscillating and stagnating plumes in the earth's lower mantle. *Earth and Planetary Science Letters* 248: 90–105.

Samuel H and Farnetani CG (2003) Thermochemical convection and helium concentrations in mantle plumes. *Earth and Planetary Science Letters* 207(1–4): 39–56.

Saxena SK, Dubrovinsky LS, Lazor P, Cerenius Y, Haggkvist P, Hanfland M, and Hu J (1996) Stability of perovskite ($MgSiO_3$) in the earth's mantle. *Science* 274: 1357–1359. http://dx.doi.org/10.1126/science.274.5291.1357.

Schubert G, Masters G, Tackley PJ, and Olson P (2004) Superplumes or plume clusters? *Physics of the Earth and Planetary Interiors* 146: 147–162.

Schuberth BSA, Bunge HP, and Ritsema J (2009) Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone? *Geochemistry, Geophysics, Geosystems* 10.

Schuberth SA, Zaroli C, and Nolet G (2012) Synthetic seismograms for a synthetic Earth: Long- period P- and S-wave traveltime variations can be explained by temperature alone. *Geophysical Journal International* 1–20. http://dx.doi.org/10.1111/j.1365-246X.2011.05333.x.

Sidorin I, Gurnis M, Helmberger DV, and Ding X (1998) Interpreting *D″* seismic structure using synthetic waveforms computed from dynamic models *Earth and Planetary Science Letters* 163: 31–41.

Simmons NA, Forte AM, and Grand SP (2006) Constraining mantle flow with seismic and geodynamic data: A joint approach. *Earth and Planetary Science Letters* 246: 109–124. http://dx.doi.org/10.1029/2006GL028009.

Simmons NA, Forte AM, and Grand SP (2007) Thermochemical structure and dynamics of the African superplume. *Geophysical Research Letters* 34: L02301. http://dx.doi.org/10.1029/2006GL028009.

Sinmyo R, Hirose K, Nishio-Hamane D, Seto Y, Fujino K, Sata N, and Ohishi Y (2008) Partitioning of iron between perovskite/postperovskite and ferropericlase in the lower mantle. *Journal of Geophysical Research* 113: B11. http://dx.doi.org/10.1029/2008JB005730.

Sleep N (1990) Hotspots and mantle plumes: Some phenomenology. *Journal of Geophysical Research* 95: 6715–6736.

Sleep NH (1988) Gradual entrainment of a chemical layer at the base of the mantle by overlying convection. *Geophysical Journal* 95: 437–447.

Solomatov VS, Olson P, and Stevenson DJ (1993) Entrainment from a bed of particles by thermal convection. *Earth and Planetary Science Letters* 120: 387–393.

Souriau A and Poupinet G (1991) A study of the outermost liquid core using differential travel times of SKS, SKKS and S3KS phases. *Physics of the Earth and Planetary Interiors* 68: 183–199.

Spera FJ, Yuen DA, and Giles G (2006) Tradeoffs in chemical and thermal variations in the post-perovskite phase transition: Mixed phase regions in the deep lower mantle? *Physics of the Earth and Planetary Interiors* 159: 234–246. http://dx.doi.org/10.1016/j.pepi.2006.07.007.

Speziale S, Zha C, Duffy TS, Hemley RJ, and Mao HK (2001) Quasi-hydrostatic compression of magnesium oxide to 52 GPa: Implications for the pressure–volume–temperature equation of state. *Journal of Geophysical Research* 106: 515–528.

Spiegelman M and McKenzie D (1987) Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth and Planetary Science Letters* 83: 137–152.

Sreenivasan B and Gubbins D (2008) Dynamos with weakly convecting outer layers: Implications for core-boundary locking. *Geophysical and Astrophysical Fluid Dynamics* 102: 395–407. http://dx.doi.org/10.1080/03091920801900047.

Sreenivasan B and Gubbins D (2011) On mantle-induced heat flow variations at the inner core boundary. *Physics of the Earth and Planetary Interiors* 187(34): 336–341.

Stacey F (1992) *Physics of the Earth*. Brookfield.

Stacey FD and Anderson OL (2001) Electrical and thermal conductivities of Fe–Ni–Si alloy under core conditions. *Physics of the Earth and Planetary Interiors* 124: 153–162.

Stacey FD and Loper DE (2007) A revised estimate of the conductivity of iron alloy at high pressure and implications for the core energy balance. *Physics of the Earth and Planetary Interiors* 161: 13–18.

Stackhouse S, Brodholt JP, Price GD, Wookey J, and Kendall JM (2005) The effect of temperature on the acoustic anisotropy of the perovskite and post-perovskite polymorphs of $MgSiO_3$. *Earth and Planetary Science Letters* 230: 1–10.

Stackhouse S, Brodholt JP, Dobson DP, and Price GD (2006) Electronic spin transitions and the seismic properties of ferrous iron-bearing $MgSiO_3$ post-perovskite. *Geophysical Research Letters* 33: L12S03.

Stanley S and Bloxham J (2004) Convective-region geometry as the cause of Uranus' and Neptune's unusual magnetic fields. *Nature* 428: 151–153. http://dx.doi.org/10.1038/nature02376.

Stanley S and Mohammadia A (2008) Effects of an outer thin stably stratified layer on planetary dynamos. *Physics of the Earth and Planetary Interiors* 168: 179–190. http://dx.doi.org/10.1016/j.pepi.2008.06.016.

Stevenson DJ (1980) Saturn's luminosity and magnetism. *Science* 208: 746–748.

Stevenson DJ (1981) Models of the Earth's core. *Science* 214: 611–619.

Stevenson DJ (1982) Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophysical and Astrophysical Fluid Dynamics* 21: 113–127.

Stevenson DJ (1987) Limits on lateral density and velocity variations in the Earth's outer core. *Geophysical Journal of the Royal Astronomical Society* 88: 311–319.

Stevenson DJ (2012) How to keep a dynamo running in spite of high thermal conductivity. In: *Abstract DI11C-03 presented at 2012 Fall Meeting AGU San Francisco, California 3–7 December*.

Stixrude L (1997) Structure and sharpness of phase transitions and mantle discontinuities. *Journal of Geophysical Research* 102: 14835–14852.

Stixrude L and Lithgow-Bertelloni C (2005) Thermodynamics of mantle minerals: 1. Physical properties. *Geophysical Journal International* 162: 610–632. http://dx.doi.org/10.1111/j.1365-246X.2005.02642.x.

Stixrude L, de Koker N, Sun N, Mookherjee M, and Karki BB (2009) Thermodynamics of silicate liquids in the deep Earth. *Earth and Planetary Science Letters* 278: 226–232. http://dx.doi.org/10.1016/j.epsl.2008.12.006.

Stolper E, Walker D, Hager BH, and Hays JF (1981) Melt segregation from partially molten source regions: the importance of melt density and source region size. *Journal of Geophysical Research* 86: 6261–6271.

Su WJ and Dziewonski AM (1997) Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle. *Physics of the Earth and Planetary Interiors* 100: 135–156.

Sun D, Song TA, and Helmberger D (2006) Complexity of *D″* in the presence of slab-debris and phase changes *Geophysical Research Letters* 33: 1207. http://dx.doi.org/10.1029/2005GL025384.

Sun D, Tan E, Helmberger D, and Gurnis M (2007a) Seismological support for the metastable superplume model, sharp features, and phase changes within the lower mantle. *Proceedings of the National Academy of Sciences* 104: 9151–9155.

Sun D, Helmberger DV, Jackson JM, Clayton RW, and Bower DJ (2013) Rolling hills on the core–mantle boundary. *Earth and Planetary Science Letters* 361: 333–342. http://dx.doi.org/10.1016/j.epsl.2012.10.027.

Sun DY, Helmberger D, Ni SD, and Bower D (2009) Direct measures of lateral velocity variation in the deep Earth. *Journal of Geophysical Research-Solid Earth* 114.

Sun XL, Song DX, Zheng SH, and Helmberger D (2007b) Evidence for a chemical-thermal structure at base of mantle from sharp lateral P-wave variations beneath Central America. *Proceedings of the National Academy of Sciences of the United States of America* 104(1): 26–30.

Suzuki A and Ohtani E (2003) Density of peridotite melts at high pressure. *Physics and Chemistry of Minerals* 30: 449–456.

Suzuki A, Ohtani E, and Kato T (1995) Flotation of diamond in mantle melt at high pressure. *Science* 269: 216–218.

Sze EKM and van der Hilst RD (2003) Core mantle boundary topography from short period PcP, PKP, and PKKP data. *Physics of the Earth and Planetary Interiors* 135(1): 27–46.

Tackley PJ (1998) Three-dimensional simulations of mantle convection with a thermo-chemical basal boundary layer: *D″*? In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region*, pp. 231–253. Washington, D.C.: American Geophysical Union.

Tackley PJ (2000a) Mantle convection and plate tectonics: Toward an integrated physical and chemical theory. *Science* 288(5473): 2002–2007.

Tackley PJ (2000b) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 1. Pseudoplastic yielding. *Geochemistry, Geophysics, Geosystems* 1. http://dx.doi.org/10.1029/2000GC000036.

Tackley PJ (2000c) Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 2. Strain weakening and asthenosphere. *Geochemistry, Geophysics, Geosystems* 1. http://dx.doi.org/10.1029/2000GC000043.

Tackley PJ (2002) Strong heterogeneity caused by deep mantle layering. *Geochemistry, Geophysics, Geosystems* 3. http://dx.doi.org/10.1029/2001GC000167.

Tackley PJ (2007) *Mantle Geochemical Geodynamics*, pp. 437–505. Elsevier.

Tackley PJ (2011) Living dead slabs in 3-D: The dynamics of compositionally-stratified slabs entering a "slab graveyard" above the core–mantle boundary. *Physics of the Earth and Planetary Interiors* 188(3–4): 150–162.

Tackley PJ (2012) Dynamics and evolution of the deep mantle resulting from thermal, chemical, phase and melting effects. *Earth-Science Reviews* 110: 1–25.

Tackley PJ, Nakagawa T, and Hernlund JW (2007) Influence of the post-perovskite transition on thermal and thermo-chemical mantle convection. In: Lay T, Yuen DA, Hirose K, and Brodholt J (eds.) *The Last Phase Transition: American Geophysical Union Monograph*.

Takafuji N, Hirose V, Mitome M, and Bando Y (2005) Solubilities of O and Si in liquid iron in equilibrium with (Mg, Fe)SiO_3 perovskite and the light elements in the core. *Geophysical Research Letters* 32: L06313. http://dx.doi.org/10.1029/2005GL022773.

Takahashi E, Shimazaki T, Tsuzaki Y, and Yoshida H (1993) Melting study of a peridotite KLB-1 to 6.5 GPa, and the origin of basaltic magmas. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 342: 105–120.

Takeuchi N (2007) Whole mantle SH velocity model constrained by waveform inversion based on three-dimensional Born kernels. *Geophysical Journal International* 169(3): 1153–1163.

Tan E and Gurnis M (2005) Metastable superplumes and mantle compressibility. *Geophysical Research Letters* 32(20).

Tan E and Gurnis M (2007) Compressible thermochemical convection and application to lower mantle structures. *Journal of Geophysical Research-Solid Earth* 112(B6).

Tan E, Leng W, Zhong S, and Gurnis M (2011) On the location of plumes and lateral movement of thermochemical structures with high bulk modulus in the 3-D compressible mantle. *Geochemistry, Geophysics, Geosystems* 12: Q07005. http://dx.doi.org/10.1029/2011GC003665.

Tanaka S (2004) Seismic detectability of anomalous structure at the top of the Earth's outer core with broadband array analysis of SmKS phases. *Physics of the Earth and Planetary Interiors* 141: 141–152.

Tanaka S (2007) Possibility of a low P-wave velocity layer in the outermost core from global SmKS waveforms. *Earth and Planetary Science Letters* 259: 486–499.

Tanaka S (2010) Constraints on the core–mantle boundary topography from P4KP–PcP differential travel times. *Journal of Geophysical Research* 115: B04310. http://dx.doi.org/10.1029/2009JB006563.

Tanaka S and Hamaguchi H (1993) Velocities and chemical stratification in the outermost core. *Journal of Geomagnetism and Geoelectricity* 45: 1287–1301.

Tarduno JA, Cottrell RD, Watkeys MK, Hofmann A, Doubrovine PV, Mamajek EE, Liu D, Sibeck DG, Neukirch LP, and Usui Y (2010) Geodynamo, solar wind, and magnetopause 3.4 to 3.45 billion years ago. *Science* 327: 1238–1240. http://dx.doi.org/10.1126/science.1183445.

Tateno S, Hirose K, Sata N, and Ohishi Y (2007) Solubility of FeO in (Mg, Fe)SiO_3 perovskite and the post-perovskite phase transition. *Physics of the Earth and Planetary Interiors* 160: 319–325. http://dx.doi.org/10.1016/j.pepi.2006.11.010.

Tateno S, Hirose K, Sata N, and Ohishi Y (2009) Determination of post-perovskite phase transition boundary up to 4400 K and implications for thermal structure in *D″* layer *Earth and Planetary Science Letters* 277: 130–136.

Thomas C, Garnero EJ, and Lay T (2004a) High-resolution imaging of lowermost mantle structure under the Cocos plate. *Journal of Geophysical Research B* 109: B08307.

Thomas C, Kendall J, and Lowman J (2004b) Lower-mantle seismic discontinuities and the thermal morphology of subducted slabs. *Earth and Planetary Science Letters* 225: 105–113.

Thomas C, Wookey J, Brodholt JP, and Fieseler T (2011) Anisotropy as a cause for polarity reversals in d″ reflections *Earth and Planetary Science Letters* 307: 369–376.

Thomas CW, Liu Q, Agee CB, Asimow PD, and Lange RA (2012) Multi-technique equation of state for Fe_2SiO_4 melt and the density of Fe-bearing silicate melts from 0–161 GPa. *Journal of Geophysical Research* 117: B10206. http://dx.doi.org/10.1029/2012JB009403.

Thompson PF and Tackley PJ (1998) Generation of mega-plumes from the core–mantle boundary in a compressible mantle with temperature-dependent viscosity. *Geophysical Research Letters* 25: 1999–2002.

Thorne M, Garnero EJ, and Grand S (2004) Geographic correlation between hot spots and deep mantle lateral shear-wave velocity gradients. *Physics of the Earth and Planetary Interiors* 146: 47–63.

Thorne MS and Garnero EJ (2004) Inferences on ultralow-velocity zone structure from a global analysis of SPdKS waves. *Journal of Geophysical Research* 109: B08301.

Thorne MS, Garnero EJ, Jahnke G, Igel H, and McNamara AK (2013a) Mega ultra low velocity zone and mantle flow. *Earth and Planetary Science Letters* 364: 59–67. http://dx.doi.org/10.1016/j.epsl.2012.12.034.

Thorne MS, Zhang Y, and Ritsema J (2013b) Evaluation of 1-D and 3-D seismic models of the Pacific lower mantle with S, SKS, and SKKS traveltimes and amplitudes. *Journal of Geophysical Research* 118: 50054. http://dx.doi.org/10.1002/jgrb.50054.

To A, Romanowicz B, Capdeville Y, and Takeuchi N (2005) 3d effects of sharp boundaries at the borders of the African and Pacific superplumes: Observation and modeling. *Earth and Planetary Science Letters* 233(1–2): 137–153.

Tolstikhin IN and Hofmann AW (2005) Early crust on top of the Earth's core. *Physics of the Earth and Planetary Interiors* 148: 109–130.

Tolstikhin IN, Kramers JD, and Hofmann AW (2006) A chemical Earth model with whole mantle convection: The importance of a core–mantle boundary layer *D″* and its early formation *Chemical Geology* 226(3–4): 79–99.

Torsvik TH, Smethurst MA, Burke K, and Steinberger B (2006) Large igneous provinces generated from the margins of the large low-velocity provinces in the deep mantle. *Geophysical Journal International* 167: 1447–1460. http://dx.doi.org/10.1111/j.1365-246X.2006.03158.x.

Torsvik TH, Smethurst MA, Burke K, and Steinberger B (2008) Long term stability in deep mantle structure: Evidence from the ca. 300 Ma Skagerrak-Centered Large Igneous Province (the SCLIP). *Earth and Planetary Science Letters* 267(3–4): 444–452.

Torsvik TH, Burke K, Steinberger B, Webb SJ, and Ashwal LD (2010) Diamonds sampled by plumes from the core–mantle boundary. *Nature* 466(7304): 352–355.

Trampert J, Vacher P, and Vlaar N (2001) Sensitivities of seismic velocities to temperature, pressure and composition in the lower mantle. *Physics of the Earth and Planetary Interiors* 124(3–4): 255–267.

Trampert J, Deschamps F, Resovsky J, and Yuen DA (2004) Probabilistic tomography maps chemical heterogenities throughout the mantle. *Science* 306: 853–856.

Tsuchiya T (2003) First-principles prediction of the *P–V–T* equation of gold and the 660-km discontinuity in Earth's mantle. *Journal of Geophysical Research* 108: 2462. http://dx.doi.org/10.1029/2003JB002446.

Tsuchiya T, Tsuchiya J, Umemoto K, and Wentzcovitch RM (2004) Phase transition in $MgSiO_3$ perovskite in the Earth's lower mantle. *Earth and Planetary Science Letters* 224: 241–248.

Tsuno K, Frost DJ, and Rubie DC (2013) Simultaneous partitioning of silicon and oxygen into the earth's core during early earth differentiation. *Geophysical Research Letters* 40: 66–71. http://dx.doi.org/10.1029/2012GL054116.

Turner JS (1986) Turbulent entrainment: the development of the entrainment assumption, and its application to geophysical flows. *Journal of Fluid Mechanics* 173: 431–471.

van der Hilst RD and Karason H (1999) Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model. *Science* 283(5409): 1885–1888.

van der Hilst RD, Widiyantoro S, and Engdahl ER (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386(6625): 578–584.

van der Hilst RD, de Hoop MV, Wang P, Shim S-H, Ma P, and Tenorio L (2007) Seismostratigraphy and thermal structure of Earth's core–mantle boundary region. *Science* 30: 1813–1817. http://dx.doi.org/10.1126/science.1137867.

Van Orman JA and Crispin KL (2010) Diffusion in oxides. *Reviews in Mineralogy and Geochemistry* 72: 757–825.

Vidale JE and Benz HM (1992) A sharp and flat section of the core–mantle boundary. *Nature* 359: 627–629.

von Bargen N and Waff HS (1986) Permeabilities, interfacial areas and curvatures of partially molten systems: Results of numerical computations of equilibrium microstructures. *Journal of Geophysical Research* 91: 9261–9276.

Vočadlo L, Alfe D, Gillan MJ, and Price GD (2003) The properties of iron under core conditions from first principles calculations. *Physics of the Earth and Planetary Interiors* 140: 101–125.

Wakabayashi D, Funamori N, Sato T, and Sekine T (2014) Equation of state for silicate melts: A comparison between static and shock compression. *Geophysical Research Letters* 41: 50–54. http://dx.doi.org/10.1002/2013GL058328.

Walker A, Forte AM, Wookey J, Nowacki A, and Kendall J-M (2011) Elastic anisotropy of *D″* predicted from global models of mantle flow *Geochemistry, Geophysics, Geosystems* 12: Q10006. http://dx.doi.org/10.1029/2011GC003732.

Walker D, Clark SM, and Cranswick LMD (2002) O_2 volumes at high pressure from $KClO_4$ decomposition: *D″* as a siderophile element pump instead of a lid on the core *Geochemistry, Geophysics, Geosystems* 3: 1070.

Wang Y and Wen LX (2004) Mapping the geometry and geographic distribution of a very low velocity province at the base of the Earth's mantle. *Journal of Geophysical Research* 109: B10305.

Wang Y and Wen LX (2007) Geometry and P and S velocity structure of the "African Anomaly" *Journal of Geophysical Research-Solid Earth* 112(B5).

Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of Earth and Planetary Sciences* 13: 201–240. http://dx.doi.org/10.1146/annurev.ea.13.050185.001221.

Wen LX (2001) Seismic evidence for a rapidly-varying compositional anomaly at the base of the Earth's mantle beneath the Indian Ocean. *Earth and Planetary Science Letters* 194: 83–95.

Wen LX (2002) An SH hybrid method and shear velocity structures in the lowermost mantle beneath the central Pacific and South Atlantic Oceans. *Journal of Geophysical Research-Solid Earth* 107(B3).

Wen LX, Silver P, James D, and Kuehnel R (2001) Seismic evidence for a thermo-chemical boundary at the base of the Earth's mantle. *Earth and Planetary Science Letters* 189(3–4): 141–153.

Whaler KA (1980) Does the whole of the Earth's core convect? *Nature* 287: 528–530.

Wicht J, Mandea M, Takahashi F, Christensen UR, Matsushima M, and Langlais B (2007) The origin of Mercury's internal magnetic field. *Space Science Reviews* 132: 261–290. http://dx.doi.org/10.1007/s11214-007-9280-5.

Wicks JK, Jackson JM, and Sturhahn W (2010) Very low sound velocities in iron-rich (Mg, Fe)O: Implications for the core–mantle boundary region. *Geophysical Research Letters* 37: L15304. http://dx.doi.org/10.1029/2010GL043689.

Williams Q (1998) The temperature contrast across *D″* In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region. American Geophysical Union Monograph*, pp. 73–81.

Williams Q and Garnero EJ (1996) Seismic evidence for partial melt at the base of the Earth's mantle. *Science* 273: 1528–1530.

Williams Q, Revenaugh JS, and Garnero EJ (1998) A correlation between ultra-low basal velocities in the mantle and hot spots. *Science* 281: 546–549.

Willis AP, Gubbins D, and Sreenivasan B (2007) Thermal core–mantle interaction: Exploring regimes for 'locked' dynamo action. *Physics of the Earth and Planetary Interiors* 165: 83–92.

Wood BJ, Walter MJ, and Wade J (2006) Accretion of the Earth and segregation of its core. *Nature* 441: 825–833. http://dx.doi.org/10.1038/nature04763.

Wookey J, Stackhouse S, Kendall J-M, Brodholt J, and Price GD (2005) Efficacy of the post-perovskite phase as an explanation of lowermost mantle seismic properties. *Nature* 438: 1004–1008.

Wright C, Muirhead KJ, and Dixon AE (1985) The P wave velocity structure near the base of the mantle. *Journal of Geophysical Research* 90: 623–634.

Wu B, Driscoll P, and Olson P (2011) A statistical boundary layer model for the mantle *D″* region *Journal of Geophysical Research* 116: B12112. http://dx.doi.org/10.1029/2011JB008511.

Wysession ME, Lay T, Revenaugh J, Williams Q, Garnero EJ, Jeanloz R, and Kellogg LH (1998) The *D″* discontinuity and its implications In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core–Mantle Boundary Region. American Geophysical Union Monograph*, pp. 273–297.

Xie S and Tackley PJ (2004a) Evolution of helium and argon isotopes in a convecting mantle. *Physics of the Earth and Planetary Interiors* 146: 417–439.

Xie S and Tackley PJ (2004b) Evolution of U–Pb and Sm–Nd systems in numerical models of mantle convection. *Journal of Geophysical Research B* 109, B11204.

Yoshida M (2008) Core–mantle boundary topography estimated from numerical simulations of instantaneous mantle flow. *Geochemistry, Geophysics, Geosystems* 9.

Zerr A, Diegeler A, and Boehler R (1998) Solidus of the Earth's deep mantle. *Science* 281: 243–245.

Zhang N and Zhong SJ (2011) Heat fluxes at the Earth's surface and core–mantle boundary since Pangea formation and their implications for the geomagnetic superchrons. *Earth and Planetary Science Letters* 306(3–4): 205–216.

Zhang N, Zhong S, Leng W, and Li Z-X (2010) A model for the evolution of the earth's mantle structure since the early paleozoic. *Journal of Geophysical Research* 115.

Zhao DP (2004) Global tomographic images of mantle plumes and subducting slabs: Insight into deep Earth dynamics. *Physics of the Earth and Planetary Interiors* 146(1–2): 3–34.

Zhong S (2006) Constraints on thermochemical convection of the mantle from plume heat flux, plume excess temperature, and upper mantle temperature. *Journal of Geophysical Research* 111: B04409. http://dx.doi.org/10.1029/2005JB003972.

Zhong S, Zuber MT, Moresi L, and Gurnis M (2000) Role of temperature-dependent viscosity and surface plates in spherical shell models of mantle convection. *Journal of Geophysical Research* 105: 11063–11082.

Zhong SJ, McNamara AK, Tan E, Moresi L, and Gurnis M (2008) A benchmark study on mantle convection in a 3-D spherical shell using CitcomS. *Geochemistry, Geophysics, Geosystems* 9.

Zhou M-F, Malpas J, Song X-Y, Robinson PT, Sun M, Kennedy AK, Lesher CM, and Keays RR (2002) *Earth and Planetary Science Letters* 196: 113122.

Zou Z, Koper K, and Cormier VF (2008) The structure of the base of the outer core inferred from seismic waves diffracted around the inner core. *Journal of Geophysical Research* 113. http://dx.doi.org/10.1029/2007JB005316.

This page intentionally left blank

7.12 Mantle Geochemical Geodynamics

PJ Tackley, Institut für Geophysik, ETH Zürich, Switzerland

© 2015 Elsevier B.V. All rights reserved.

7.12.1 Observations and the Origin of Heterogeneity 522
7.12.1.1 Introduction 522
7.12.1.2 Bulk Earth and BSE models 522
7.12.1.3 Fractionation of Major and Trace Elements 523
7.12.1.4 Trace-Element Budgets of the Crust and Mantle 523
7.12.1.4.1 Reservoir of 'missing' elements 524
7.12.1.5 Origin of Heterogeneities 524
7.12.1.5.1 Primordial differentiation (fractionation of magma ocean, core formation, and early atmosphere) 524
7.12.1.5.2 Formation of continental and oceanic crust 524
7.12.1.5.3 Processing rate of mid-ocean ridge melting and residence time 525
7.12.1.5.4 Reactions with the core 526
7.12.1.6 MORB and OIB Trace-Element Signatures 527
7.12.1.6.1 Noble gases 527
7.12.1.6.2 The end-member 'zoo' 529
7.12.1.6.3 Nd isotopes and early differentiation and mixing processes 530
7.12.1.7 Length Scales: Geochemical and Seismological 530
7.12.1.7.1 Geochemical scales 530
7.12.1.7.2 Seismological scales 531
7.12.1.8 Geophysical Constraints in Favor of Whole-Mantle Convection 531
7.12.1.9 Overview of Geochemical Mantle Models 532
7.12.1.9.1 Vertical layering 532
7.12.1.9.2 Distributed heterogeneity 534
7.12.1.10 Mantle–Exosphere (Biosphere) Interactions: Water and Carbonate Cycling 534
7.12.2 Mantle Stirring and Mixing of Passive Heterogeneities 534
7.12.2.1 Stirring, Stretching, Mixing, and Dispersal 534
7.12.2.2 Stretching Theory: Laminar and Turbulent Regimes 535
7.12.2.2.1 Finite deformation tensor 535
7.12.2.2.2 Asymptotic stretching rate 535
7.12.2.2.3 Measures of stretching 536
7.12.2.3 Stretching: Numerical Results 537
7.12.2.3.1 General discussion 537
7.12.2.3.2 Steady-state 2-D flows 537
7.12.2.3.3 Time-dependent 2-D flows 537
7.12.2.3.4 3-D flows 538
7.12.2.3.5 Effect of viscosity variations 539
7.12.2.4 Dispersal 540
7.12.2.4.1 Measures of dispersal 540
7.12.2.4.2 Numerical results 541
7.12.2.4.3 Eddy diffusivity? 542
7.12.2.5 Residence Time 543
7.12.2.5.1 Effect of chemical density differences 543
7.12.2.5.2 Effect of viscosity variations and convective vigor 544
7.12.2.6 Mixing Time 544
7.12.2.6.1 Background 544
7.12.2.6.2 Laminar flows 545
7.12.2.6.3 Turbulent regime 545
7.12.2.6.4 Summary of different 'mixing time' estimates 546
7.12.2.7 Spectrum of Chemical Heterogeneity 546
7.12.3 Mantle Convection with Active (Buoyant) Chemical Heterogeneity 547
7.12.3.1 Stability and Dynamics of Chemical Layering in a Convecting Mantle 547
7.12.3.1.1 The balance between chemical and thermal buoyancy 547
7.12.3.1.2 Initially stable layering: Pattern and dynamics 548
7.12.3.1.3 Influence of a dense layer on plume dynamics 550
7.12.3.1.4 Entrainment of surrounding material by mantle plumes 551

7.12.3.1.5 Ultra-low-velocity zones 552
7.12.3.2 Long-Term Entrainment of Stable Layers 552
7.12.3.2.1 Comparison of entrainment laws and estimates 552
7.12.3.3 Evolution of a Mantle with Ongoing Differentiation 555
7.12.3.3.1 General evolution 555
7.12.3.3.2 Magmatic heat transport 556
7.12.3.3.3 Development of a dense layer by gravitational settling of subducted crust 556
7.12.3.3.4 Chemical layering induced by phase transitions 559
7.12.3.3.5 The effect of chemical layering on core heat flow and planetary thermal evolution 561
7.12.3.4 On the accuracy of numerical thermochemical convection calculations 562
7.12.4 Convection Models Tracking Trace-Element Evolution 562
7.12.4.1 Noble Gases 562
7.12.4.1.1 Helium and argon 562
7.12.4.1.2 Origin of $^{3}He/^{4}He$ 563
7.12.4.2 Other Isotopes (U, Pb, Nd, and Sm) and the Origin of HIMU 564
7.12.4.3 Extension to Include Continental Differentiation 565
7.12.4.4 Isotopic Variations at Hotspots and Transport by Plumes 565
7.12.5 New Concepts and Future Outlook 566
7.12.5.1 Revisiting Noble Gas Constraints 566
7.12.5.1.1 Ar and He budget 566
7.12.5.1.2 High $^{3}He/^{4}He$ 566
7.12.5.2 Improved Recipes for Marble Cakes and Plum Puddings 567
7.12.5.2.1 Two recipes 567
7.12.5.2.2 Two-stage melting 567
7.12.5.2.3 Source statistics and component fractions 568
7.12.5.2.4 Melting of a heterogeneous source 568
7.12.5.2.5 Can OIB and MORB be produced by the same statistical assemblage? 569
7.12.5.2.6 Basal Magma Ocean 570
7.12.5.2.7 Internal differentiation: 'Upside-down' and 'hybrid pyroxenite' 571
7.12.5.3 Water: Influence on Convection and the Transition Zone Water Filter 571
7.12.5.4 Outlook and Future Directions 572
7.12.5.4.1 Can geochemical and geophysical observations be reconciled with current paradigms? 572
7.12.5.4.2 Quantitative modeling approaches 573
Acknowledgments 573
References 573

7.12.1 Observations and the Origin of Heterogeneity

7.12.1.1 Introduction

Geochemical constraints on mantle heterogeneity are thoroughly reviewed in several papers (e.g., Caro, 2011; Graham, 2002; Hilton and Porcelli, 2003; Hofmann, 2003; White, 2010), but as they are so important for understanding this chapter, it is appropriate to briefly summarize here the basic observations and 'traditional' interpretations. New interpretations are discussed in **Section 7.12.5**. It is noted that with modern analysis techniques, particularly MC-ICPMS machines, the number of isotopes that geochemists are measuring has increased, and while many isotope systems are consistent in their behavior, additional complexity in interpretation is introduced by isotope systems that are affected by additional processes.

This review starts with models of the bulk silicate Earth (BSE) composition and then constraints on the variations of composition, which fall into three main categories: (i) mass balance arguments based on the inferred composition of the mid-ocean ridge basalt (MORB) source and continental crust, (ii) noble gas constraints, and (iii) end-member components identified by analysis of the trace-element isotopic compositions of ocean island basalts (OIBs) and MORBs. This section also reviews geophysical constraints on chemical heterogeneity and mantle structure and the major types of mantle conceptual model that have been proposed to explain geochemical and/or geophysical observations.

7.12.1.2 Bulk Earth and BSE models

The bulk composition of the whole Earth (i.e., core + mantle + crust) is estimated using a variety of constraints as reviewed in Palme and O'Neill (2003), particularly the composition of the Sun, meteorites, and upper mantle-derived rocks. The relative abundance of refractory elements is typically thought to be chondritic (i.e., solar), while nonrefractory to volatile elements show increasing degrees of depletion. The composition of the silicate part, that is, Bulk Silicate Earth (BSE), is further influenced by core extraction, which enriched the mantle in nonsiderophile elements but depleted it in siderophile elements. Knowledge of the trace-element content of the BSE is important in making 'mass balance' arguments summarized

later, as is the knowledge of the composition of the continental crust (e.g., Rudnick, 1995; Rudnick and Fountain, 1995; Taylor and McLennan, 1995). A commonly used compositional BSE model is that of McDonough and Sun (1995), while a more depleted model was proposed by Lyubetskaya and Korenaga (2007a). It is also possible that the Earth is not chondritic, as argued, for example, by Drake and Righter (2002) and Campbell and O'Neill (2012). The reader is referred to the previously cited literature for further detailed discussions.

7.12.1.3 Fractionation of Major and Trace Elements

When solid and liquid coexist in equilibrium with each other in a multicomponent material such as mantle silicates, the composition of the solid and liquid components is different in both major and trace elements. The two main situations for solid–liquid equilibration are partial melting (caused by, e.g., a decrease of pressure in an upwelling or by infiltration of volatiles) and the crystallization of a magma ocean.

A crude way of viewing major-element fractionation is that different minerals have different melting temperatures – for example, in the shallow mantle, garnet and clinopyroxene melt at a lower temperature than olivine. Thus, partial melting followed by melt segregation and eruption causes both the crust that forms from the resulting melt and the residual solid to have a different composition from the original rock. In a cooling magma ocean, this results in the solid crystals that form having a different composition from the remaining liquid, which might allow them to settle to the top or bottom due to related density differences. Indeed, density differences that result from differences in major-element composition may well play an important role in mantle dynamics, as discussed in later sections.

Trace-element partitioning between solid and liquid is expressed in terms of partition coefficients d, which describe the equilibrium concentration of the element in the solid divided by the concentration in the liquid:

$$d = \frac{C_{\text{solid}}}{C_{\text{liquid}}} \quad [1]$$

where C_{solid} and C_{liquid} are the concentrations of the element in the solid and liquid, respectively. A partition coefficient $\ll 1$ means that the element preferentially enters the melt and is known as an incompatible element. Conversely, a partition coefficient $\gg 1$ indicates that the element is compatible and stays in the solid. The partition coefficient depends on the element (because different elements have different sizes and electronic properties), which results in trace elements being fractionated during melting, that is, the ratio of their concentrations changes. Partition coefficients may also depend on pressure (depth), oxygen fugacity, and the minerals present. Different isotopes of the same element have partition coefficients that are almost the same and can be treated as identical for the situations discussed in this chapter. For more information, see Righter and Drake (2003), Wood and Blundy (2003), and **Chapter 2.04**.

This fractionation of trace elements plays a key role in the evolution of isotope ratios of interest in mantle geochemistry. In a simple radioactive decay system in which a parent isotope radioactively decays to a daughter that is a different element, the parent to daughter ratio (P/D) will change as a result of melting and crustal formation, whereas the ratio of the daughter to a stable isotope of the same element (D/D_S) will not change. This change in (P/D) affects the future rate of change of (D/D_S). By analyzing the (D/D_S) ratio in two or more samples that have experienced fractionation in a single event, it is possible to derive the elapsed time since the fractionation event, and this is the principle of radiogenic dating. The specific example of Pb–Pb dating is more complicated because it combines two decay systems, but this also makes it more robust (for details, see, e.g., Stacey, 1992).

The concentration of trace elements in the melt depends not only on the partition coefficient but also on the melt fraction. For example, a highly incompatible element essentially enters the melt, which means that its concentration in the melt is inversely proportional to melt fraction. In the more general case, it is straightforward to derive an equation for the equilibrium concentration of a trace element in the melt (Shaw, 1970):

$$C_{\text{liquid}} = \frac{C_0}{F + d(1-F)} \quad [2]$$

where F is the melt fraction and C_0 is the concentration in the total (bulk) system. Thus, the enrichment of a magmatic product in trace elements relative to its source can be used to estimate the melt fraction.

7.12.1.4 Trace-Element Budgets of the Crust and Mantle

The trace-element composition of the continental crust and the inferred composition of the MORB source are often compared to the composition of the BSE discussed in **Section 7.12.1.2**. This comparison indicates that the continental crust is strongly enriched in incompatible trace elements – by a factor of ~50–100 in the most incompatible elements, whereas the MORB source is depleted. Furthermore, the patterns of enrichment and depletion (i.e., as a function of elements) appear to be complementary. The enriched continental crust and the depleted MORB source region have thus long been thought to be complementary in their trace-element signature, implying that the MORB source region is the residue from melting that produced the continental crust (e.g., Hofmann, 1988). Mass balance calculations have been used to estimate the fraction of the mantle that must have been depleted to make the continental crust (e.g., Allegre et al., 1983a; Davies, 1981; Hofmann, 1988; Jacobsen and Wasserburg, 1979; O'Nions et al., 1979) with the estimated depleted reservoir size ranging greatly between 30% and 97% of the mantle (50–80% of the mantle according to Hofmann (2003)) and the rest of the mantle being primitive, that is, roughly chondritic. A curious feature of this calculation is that the degree of melting required to produce continental crust is small, for example, ~1% (Hofmann, 2003) – about an order of magnitude lower than the 8–10% melting thought to produce present-day MORB. This indicates that incompatible trace elements must be scavenged from a large volume and concentrated into a small volume in order to produce the concentrations observed in continental crust, as discussed in **Section 7.12.1.5.2**.

7.12.1.4.1 Reservoir of 'missing' elements

The previous mass balance calculation implies the existence of a primordial reservoir of primitive (i.e., BSE) composition containing the trace elements that are 'missing' from the continental crust plus depleted (MORB source) mantle. This inferred reservoir of volume 20–50% of the mantle is often argued to be important in explaining the mantle's heat budget, as it would contain a large amount of heat-producing elements, and the concentration of heat-producing elements in the MORB source appears to be too low to explain the mantle's heat budget. The reservoir is argued to be 'hidden,' in the sense that it produces no observed geochemical signature, except perhaps that of high $^{3}He/^{4}He$. As an alternative to this relatively large reservoir, it has been proposed that the 'missing' trace elements are contained in a small, early enriched reservoir (EER) (e.g., Boyet and Carlson, 2006; Coltice and Ricard, 1999; Tolstikhin and Hofmann, 2005; Tolstikhin et al., 2006). Another possibility is that the bulk Earth composition is more depleted than previous estimates (Lyubetskaya and Korenaga, 2007a), which would not require any hidden reservoir (Lyubetskaya and Korenaga, 2007b).

7.12.1.5 Origin of Heterogeneities

Chemical variations inside the mantle could arise either from primordial differentiation of the Earth, through subsequent magmatism, or through possible chemical reactions with the core; these are discussed in this section. Quantitative models of either process require knowledge of the bulk composition of the mantle, which was discussed in **Section 7.12.1.2**.

7.12.1.5.1 Primordial differentiation (fractionation of magma ocean, core formation, and early atmosphere)

The extent to which the mantle might have started off layered due to fractional crystallization of a magma ocean has been hotly debated over the years. Key issues are whether a deep magma ocean existed, and if so whether its cooling would result in a fractionation due to crystal settling, or whether it would instead solidify into a chemically homogeneous mantle. This depends on various factors such as the type of transient atmosphere present, grain size, viscosities, and convection mode (Abe, 1997; Elkins-Tanton et al., 2003; Solomatov and Stevenson, 1993a,b). The existence of a deep magma ocean has previously been argued against because the geochemical signature of the mantle is inconsistent with wholesale fractionation (e.g., Ringwood, 1990a), but if magma ocean convection is vigorous enough that crystals do not settle (Solomatov et al., 1993; Solomatov and Stevenson, 1993b; Tonks and Melosh, 1990), then this difficulty vanishes. Furthermore, models of the Earth's formation predict collisions with planetesimals the size of the Moon or even Mars (the latter of which offers the most popular explanation for the formation of the Moon), which would inevitably result in large-scale melting and a deep terrestrial magma ocean (e.g., Benz and Cameron, 1990; Melosh, 1990). Another possibly important aspect of magma ocean evolution is its interaction with the primordial atmosphere, through which it might have absorbed volatile elements (e.g., Ballentine, 2002; Dasgupta, 2013; Elkins-Tanton, 2008; Zahnle et al., 2010), although Ballentine et al. (2005) suggested that this was not important for primitive volatile (particularly He/Ne) acquisition.

While there is therefore much uncertainty about magma ocean evolution, a synthesis by Solomatov (2000) offers a scenario that is consistent with both geochemical and dynamic constraints. In this case, following a giant impact, the lower mantle solidifies rapidly (i.e., <1000 years) and is undifferentiated, leaving a shallow magma ocean in which fractionation and much slower cooling occur. This is consistent with the earlier work of Abe (1997), who also found that following a giant impact, lower mantle differentiation is uncertain, whereas upper mantle differentiation is likely, and a shallow magma ocean could remain for 100–200 My.

Another chemically important process is the separation of the metal core from the remaining silicate. Metal–silicate separation is likely to have occurred in (at least) two environments (Karato and Murthy, 1997; Stevenson, 1990): small metal droplets (arising either from dispersed metal particles in undifferentiated solid chondritic material or from emulsification of the cores of differentiated impacting bodies) falling through a magma ocean, where separation would occur rapidly (Hoink et al., 2006; Rubie et al., 2003; Solomatov, 2000; Tonks and Melosh, 1990; Ulvrova et al., 2011), and large metal diapirs (arising either from pools of liquid metal at the base of a shallow magma ocean or from the cores of impacting differentiated bodies) sinking through solid silicate or primitive material. The details of these strongly influence the partitioning of trace elements between the mantle and the core as well as the extent to which the core and mantle are chemically equilibrated in general. Specifically, chemical equilibration is efficient in a magma ocean but has generally been thought to be less effective for large sinking diapirs, leading to the view that metal–silicate equilibration last happened at temperatures and pressures relevant to the base of the magma ocean (Murthy and Karato, 1997). However, Samuel and Tackley (2008) showed that sinking iron diapirs of radius 1–100 m can chemically equilibrate, so this conclusion is called into question. Such mechanisms are relevant to questions such as whether the core contains potassium, carbon, or noble gases, as well as whether reactions between the mantle and the core take place.

In summary, at present, the consensus from geochemical constraints combined with dynamic considerations is that primordial chemical stratification in the deep mantle is unlikely, whereas primordial upper mantle chemical stratification is likely. Partitioning of elements between the mantle and the core is sensitive to poorly known processes that occurred during core–mantle separation.

7.12.1.5.2 Formation of continental and oceanic crust

It is clear that much differentiation has taken place due to melting subsequent to Earth's formation. The most obvious manifestations of this are (i) the formation of the continental crust, which is strongly enriched (by a factor of up to 50–100) in incompatible trace elements compared to BSE models and can survive over billions of years, and (ii) the formation of the oceanic crust, which is less enriched and gets subducted back into the mantle after 100–200 My, possibly to get mixed again (see later discussion). A thorough review of crustal formation and its geochemical consequences was given in Hofmann

(2003); here, the main points relevant to the discussion in this chapter are summarized.

7.12.1.5.2.1 Continental crust

As discussed in the preceding text, from a trace-element perspective, the (enriched) continental crust is typically thought to be complementary to the depleted mantle, with an implied degree of melting of ~1%. This is very low and indicates that the continental crust is produced by more complex processes than simple partial melting (by ~1%) and eruption – indeed, melting experiments carried out on peridotite produce magmas that do not resemble (i.e., are less 'evolved' than) the continental crust in their major-element composition (Hirose and Kushiro, 1993; Rudnick, 1995). The continental crust has two layers, a lower crust with an approximately basaltic composition and an upper crust with a roughly granodioritic composition. A two-step process is commonly envisaged (Hawkesworth and Kemp, 2006). First, a new juvenile continental crust with a basaltic composition is produced mostly by water-induced melting of the mantle above subducting slabs at convergent margins; Stern and Scholl (2010) estimated 1.5 $km^3 year^{-1}$ from oceanic arcs and 1.0 $km^3 year^{-1}$ from continental arcs out of a total of 3.2 $km^3 year^{-1}$. This water-induced melting is caused by dehydration of subducted crust and movement of the resulting fluid through the mantle wedge (together with incompatible trace elements), followed by metasomatism and melting. A secondary source of new continental crust is intraplate volcanism (i.e., flood basalts on continents or accretion of oceanic plateaus). The lower continental crust is thought to have a similar composition to the juvenile continental crust (Hawkesworth and Kemp, 2006). Second, the upper continental crust forms by partial melting and resulting differentiation of this juvenile crust. The degree of partial melting is about 14% (Hawkesworth and Kemp, 2006), which results in a large volume of residue that is not observed in continents; this residue is dense so is thought to delaminate and founder into the mantle, where it might contribute to OIB sources. Delamination of lower crustal residual material is thus an important process, although it may only be possible over a certain range of temperature and pressure conditions (Jull and Kelemen, 2001). Indeed, Tatsumi et al. (2013) found that the residue is denser than pyrolite all the way to the core–mantle boundary (CMB) and may therefore contribute to any dense piles that exist there. Processes generating continental crust may have been different in the Archean and Hadean (e.g., Condie and Kröner, 2013; Foley et al., 2003; Rudnick, 1995; Taylor and McLennan, 1995). Johnson et al. (2014) calculated that the thickness of very early crust was limited to about 45 km by phase transitions to denser minerals; below this depth, the relatively dense crust drips back into the mantle.

The growth history of continents is highly uncertain and reflects a balance between processes of production of new continental material and the loss of old continental material from subduction, at collision zones or by foundering of lower crustal material. Several continental growth curves have proposed, largely on the basis of geochemical arguments; for a comparison of growth curves and review, see Harrison (2009) and Taylor and McLennan (1995), and for a recent model constrained by isotopes in zircons, see Dhuime et al. (2012). Continents may have grown in several pulses such as at 2.7, 1.9, and 1.2 Ga ago (Hawkesworth and Kemp, 2006). At the present day, creation and destruction of continental crust are roughly balanced (Stern and Scholl, 2010). Very little ancient continental crust is preserved; however, recent zircon data imply that continental crust already existed 4.4–4.5 Gy before present and was metamorphosed under aqueous conditions at pressures corresponding to >20 km depth and moderate temperatures (Harrison, 2009; Harrison et al., 2005).

Also important and compositionally distinct is the continental mantle lithosphere. A useful review of its composition, formation, and destruction of this can be found in Lee et al. (2011). Formation mechanisms and the compositions of the continental lithosphere and crust have changed with time due to higher mantle temperatures in the Archean (Herzberg and Rudnick, 2012).

7.12.1.5.2.2 Oceanic crust

While continental crust formation is generally considered to be the dominant influence on the trace-element abundances in the mantle (the present-day oceanic crust is much less enriched and smaller in volume), a large fraction of the mantle has passed through the melting ('processing') zone beneath mid-ocean ridges, and thus, the production and recycling of oceanic crust have had a major influence on mantle heterogeneity. The rate at which this happens is somewhat uncertain and is discussed in detail in the next section. Regardless of the exact rate of processing, subducted MORB is clearly a volumetrically significant component in the mantle that is responsible for much of the observed isotopic heterogeneity (e.g., Stracke, 2012) so should not be ignored in calculating mantle geochemical evolution.

7.12.1.5.3 Processing rate of mid-ocean ridge melting and residence time

The volume of material that has been processed by MOR melting is quite uncertain, because of differing estimates of the present-day rate of processing and uncertainty about how this changed in the past over billions of years. One measure of the processing rate is the 'residence time,' which for random sampling of the mantle is equal to the length of time taken for the mass of the mantle to pass through the processing zone. At present rates of seafloor spreading, estimates of this processing time or residence time have varied widely, from 2.9 to 10 Ga. Example estimates are 4 Gy (O'Connell and Hager, EOS 61, 373, 1980, as referenced in Davies (1981)), 10 Gy (Gurnis and Davies, 1986b), 5.7 Gy (Kellogg and Wasserburg, 1990), 9.5 Gy (Phipps Morgan, 1998), 3–6 Gy (Davies, 2002), 4.5 Gy (Xie and Tackley, 2004b), and 2.9 Gy (Coltice, 2005; Huang and Davies, 2007c). It is worth analyzing why there is such a large variation in these estimates. The processing rate, expressed as the volume of mantle processed per year, is given by

$$P = Ad\rho \qquad [3]$$

where A is the area of oceanic seafloor produced per year, d is the depth of onset of melting beneath spreading centers, and ρ is the density of the processed material. Observations of the age distribution of ocean floor (e.g., summarized in Phipps Morgan (1998)) indicate that $A = 2.7\ km^2 year^{-1}$ for seafloor

produced in the last 1 Ma, or $A = 2.9$ km^2 year^{-1} based on an average over the last 100 Ma, the latter being similar to the 3.0 km^2 year^{-1} estimated by Stacey (1992). Other authors have preferred to use 'back of the envelope' estimates, for example, Davies (2002) assumed a ridge length of 40 000 km and spreading velocity of 5–10 cm year^{-1} (giving $A = 2$–4 km^2 year^{-1}) and Coltice (2005) assumed a ridge length of 60 000 km and plate velocity of 5 cm year^{-1} ($A = 6$ km^2 year^{-1}). Regarding the onset depth of melting d, Phipps Morgan (1998) assumed that the oceanic crustal thickness of 6.5 km was produced with 10% melt ($d = 65\rho_{crust}/\rho_{mantle}$ km), whereas Davies (2002) assumed that $d = 100$ km, and Coltice (2005) assumed that $d = 150$ km. Using the observationally constrained $A = 2.9$ km year^{-1} together with, for example, $d = 80$ km and peridotite density $= 3000$ kg m^{-3} leads to $P = 6.96 \times 10^{14}$ kg year^{-1}, which gives a residence time of 5.74 Gy to process the mantle mass of 4×10^{24} kg. Thus, estimates of processing (residence) time as low as 3 Gy should be regarded as extreme lower limits. Estimates of the present-day processing time based on geochemical fluxes (Albarède, 2005b) range from 4 to 9 Gy, bracketing the geophysical value discussed earlier; this estimate is further discussed in **Section 7.12.2.5**.

It is likely that the rate of seafloor spreading and hence crustal production was considerably higher in the past when the mantle was hotter, which would greatly reduce the processing/residence time in the past (Gurnis and Davies, 1986b). In addition, the depth of melting would have been larger due to the hotter mantle, as considered by Davies (2002) and Tajika and Matsui (1992, 1993), further increasing the processing rate. The appropriate increase of seafloor spreading in the past is not known: in simple, internally heated convection in secular equilibrium, the surface velocity scales as (radiogenic heat production). Davies (1980) and Phipps Morgan (1998) gave detailed arguments why this also applies to plate tectonics. It is, however, likely that the early Earth did not have plate tectonics as it exists today because the thicker oceanic crust and younger oceanic plates (when they reach subduction zones) would make plates difficult to subduct (Davies, 1990, 1992; Vlaar, 1985) and perhaps because the mechanical oceanic lithosphere was thicker because water depletion, which causes an increase in viscosity, occurred deeper (Korenaga, 2003). Thus, the most likely increase in crustal production rate lies between constant and H^2, where H = radiogenic heat production rate.

To give some quantitative examples of the influence of faster processing in the past, (i) Xie and Tackley (2004b) calculated the volume of oceanic crust produced over the age of the Earth to be ~10% of the mantle mass if the present-day production rate is applied to the whole Earth's history, 54% if the H^2 law of Davies (1980) is applied, or 82% if the increase in depth of melting due to higher mantle temperatures is additionally taken into account. In the latter two cases, this clearly implies that the same material melts two or more times, as it is not possible to produce this amount of MORB from the residue of continental crust production. (ii) Davies (2002) calculated that the H^2 law leads to four times the amount of processing over geologic time, which he expressed as being equivalent to 18 Gy of constant-rate processing. This means that 98.6% of the mantle is processed if the present-day residence/processing time is 4.2 Gy.

Outgassing of ^{40}Ar is also closely related to the processing/residence time. As mentioned earlier, the ~50% outgassing constraint can be matched by a residence time of 9.5 Gy and a processing rate scaling with H^2 (Phipps Morgan, 1998). With a much smaller processing time varying from 1.98 to 4.23 Gy through the calculation, Huang and Davies (2007a) obtained 81% argon outgassing, suggesting that such processing times are too small.

These issues are further discussed from the perspective of modeling in **Section 7.12.2.5**.

7.12.1.5.4 Reactions with the core

Some researchers have studied the possibility of an iron-rich layer forming above the CMB caused by reactions with the core. There is some geochemical evidence (based on the Os–Pt system) for mass exchange between the mantle and the core (Brandon et al., 1998, 2003; Humayun et al., 2004; Walker et al., 1995), but this is controversial (Brandon and Walker, 2005; Schersten et al., 2004).

Experiments indicate that iron reacts with silicate perovskite to form iron oxide and silica (Knittle and Jeanloz, 1989, 1991). Experiments with liquid iron and solid silicate show that the silicate becomes depleted in Fe and O (Goarant et al., 1992). Liquid iron also reacts with Al_2O_3 at CMB pressures (Dubrovinsky et al., 2001) but not with silica (Dubrovinsky et al., 2003). Using thermodynamic calculations, Song and Ahrens (1994) determined that the observed reactions are thermodynamically possible. Thus, core–mantle reactions are expected from an experimental standpoint, although it is possible that the presence of water inhibits them (Boehler et al., 1995; Poirier et al., 1998).

It appears difficult to generate significant mantle heterogeneity this way, because the thickness of the reaction zone is relatively small. Poirier (1993) and Poirier and Le Mouel (1992) found that liquid Fe can penetrate 1–100 m into the mantle along grain boundaries if it reacts with the silicate. Knittle and Jeanloz (1991) proposed that the Fe ascents by a capillary action driven by surface tension. However, recently, based on laboratory experiments, Otsuka and Karato (2012) identified a morphological instability by which Fe-rich liquids could penetrate into (Mg, Fe)O, due to the two not being in chemical equilibrium. They estimated that via this mechanism, Fe could penetrate ~50 km into the mantle in a billion years, which is much faster than the previous estimates. A dynamic mechanism could be a suction mechanism in areas of downwelling (where the CMB is depressed), which may entrain Fe of order ~1 km into the mantle with fractions of up to 10%, depending on parameters such as the effective bulk viscosity (Kanda and Stevenson, 2006; Stevenson, 1988). Petford et al. (2005) also proposed a dynamically assisted Fe transfer mechanism assuming that the D″ region behaves as a poro-viscoelastic granular material with dilatant properties, but the required strain rates are rather high (10^{-12} s^{-1}). Thus, it appears that Fe might be able to infiltrate up to ~1–10s km at significant volume fractions, but this needs further study.

If Fe infiltrates the mantle in some regions in a thin layer and reacts to form dense products, the question is then whether the continual flux of new mantle material through this reaction zone is sufficient, over geologic time, to affect a volumetrically significant amount of mantle material. Knittle

and Jeanloz (1991) suggested that the material would not be raised far into the mantle because of its increased density and could accumulate into a approximately few 100 km thick D″ layer, but this needs to be tested using detailed modeling. Numerical models of this performed to date have parameterized core reactions as an effective chemical diffusivity, which may not correctly capture the process. The most advanced model is that of Kellogg (1997), who parameterized the infiltration of Fe using a chemical diffusivity 100–1000 times smaller than the thermal diffusivity and found that large density contrasts ($B=10$) result in a thin (<100 km) layer that slows further growth, whereas lower density contrasts ($B=1$) result in an undulating layer several 100s km thick that piles up under upwellings, as is observed in earlier calculations (Hansen and Yuen, 1988). If a metal-bearing layer does build up, it could have a thermal conductivity an order of magnitude higher than that of silicates (Manga and Jeanloz, 1996), which could further influence mantle convection.

It has been proposed that silicate melt (partial melt fractions up to 30%) exists in some areas at the CMB as an explanation of seismically observed ultra-low velocity zones (ULVZs) (Lay et al., 2004; Williams and Garnero, 1996), as further discussed in **Section 7.12.3.1.5**. Such areas of partial melt could have interesting geochemical consequences, as discussed in Labrosse et al. (2007) and in **Section 7.12.5.2.5**. In the present context, they might facilitate reactions between the mantle and the core due to the larger flux of mantle material that comes into contact with the core and also because chemical diffusion is faster in liquids (see also **Chapter 7.11**).

7.12.1.6 MORB and OIB Trace-Element Signatures

7.12.1.6.1 Noble gases

The most common noble gases considered in mantle geochemistry are the isotopes of He and Ar, with Ne and Xe also sometimes considered.

7.12.1.6.1.1 ^{3}He/^{4}He ratios

The two isotopes of helium are ^{3}He, which is primordial, and ^{4}He, which in rocks is almost all produced by the radioactive decay of U and Th (the primordial abundance of ^{4}He is small compared to that of U and Th because He is volatile). Histograms of the ratio of ^{3}He/^{4}He for MORBs and OIB show clear differences. MORBs have a relatively constant ^{3}He/^{4}He ratio with a mean value of 1.16×10^{-5}, which is a factor of 8.4 higher than the atmospheric value (Allegre et al., 1995; Graham, 2002; Kurz et al., 1983). The range of ^{3}He/^{4}He ratios between 7 and 9 is thus taken as representative of the depleted upper mantle. OIB ^{3}He/^{4}He ratios are, in contrast, extremely heterogeneous, extending to both higher and lower values than MORB, with the highest ratios measured (at Baffin Island) about 50 times the atmospheric value (Stuart et al., 2003).

There are several possible explanations for the high ^{3}He/^{4}He material. The 'traditional' interpretation is that this component comes from an undegassed, 'primitive' source, because noble gases are highly incompatible and volatile and are thus expected to enter any melt and be heavily outgassed prior to or during eruption (Allegre et al., 1983b; Farley et al., 1992; Porcelli and Wasserburg, 1995; note that outgassing requires CO_2 bubbling, which occurs at depths <60 km). This component has traditionally been identified as the lower mantle, consistent with the need to keep it intact for billions of years, with other arguments for the lower mantle being primitive, and with the idea of plumes sampling the lower mantle and spreading centers sampling the upper mantle. However, the actual volume and location of this component are not constrained. If the 'primitive' interpretation is correct, one might expect the absolute concentration of He in OIB to be higher than that in MORB, but this is not observed, perhaps because magma chamber processes cause the loss of He (Farley and Neroda, 1998).

The leading alternative explanation for high ^{3}He/^{4}He is that it is associated with recycled depleted oceanic lithosphere (Albarède, 1998; Anderson, 1998; Coltice and Ricard, 1999, 2002; Ferrachat and Ricard, 2001). This is based on the fact that high ^{3}He/^{4}He simply requires high ^{3}He/(U+Th) because decay of U and Th produces ^{4}He, therefore, an alternative to high [^{3}He] is low [U+Th], which is found in material that has been depleted in incompatible trace elements due to melting. The evolution of He ratios is illustrated in **Figure 1**. Note that ^{3}He/^{4}He decreases with time in all 'reservoirs' due to the radiogenic ingrowth of ^{4}He. If the ratio ^{3}He/(U+Th) is changed in different components due to fractionation (during which ^{3}He/^{4}He remains the same), then with time, it will evolve ^{3}He/^{4}He that is higher or lower than the original material. If He is less incompatible than U and Th (e.g., Graham et al., 1990), and as found in some laboratory experiments (Parman et al., 2005) although not others (Heber et al., 2007), then ^{3}He/(U+Th) is higher in the depleted residue, and after some time, its ^{3}He/^{4}He will be higher than that of the original material. However, Anderson (1998), Coltice and Ricard (1999, 2002), and Ferrachat and Ricard (2001) showed that

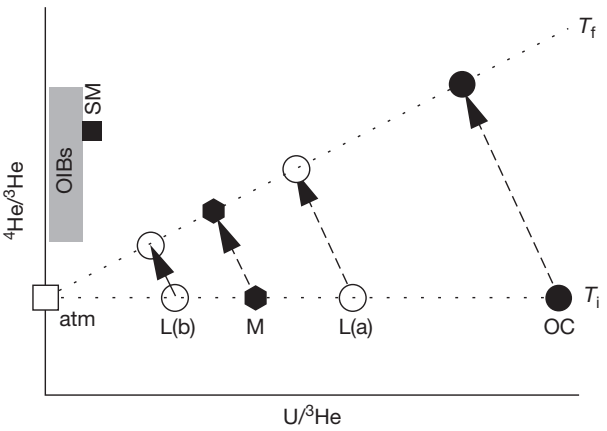


Figure 1 Schematic diagram showing the helium isotope evolution of different mantle components, from Coltice and Ricard (2002). M, mantle; L, lithosphere; atm, atmosphere; OC, oceanic crust. When M melts, it produces oceanic crust and lithosphere (a) or (b) depending on the relative partition coefficients of U and He, and some He is lost to the atmosphere. Between melting events, U decays and produces ^{4}He, so U/^{3}He decreases and ^{4}He/^{3}He increases. The shallow mantle (SM) is a mixture of recycled lithosphere and oceanic crust, so recycled lithosphere will always have lower ^{4}He/^{3}He (high ^{3}He/^{4}He) than the average, regardless of the relative partition coefficients. With successive melting events, the mantle average ratios move to the right due to He outgassing.

it is not necessary for He to be less incompatible if the residuum is stored for substantial time periods while the rest of the mantle becomes increasingly outgassed. To understand this, note that in the crust, $^{3}He/(U+Th)$ is extremely low due to degassing of He and concentration of U and Th, so the crust will have much lower $^{3}He/^{4}He$ regardless of the exact relative He and U partitioning. If the mantle is almost all differentiated, such that to first order it is made of a mixture of subducted crust and depleted residue, then the average mantle $^{3}He/^{4}He$ will inevitably be in between the ratios in residue and subducted crust, making residue $^{3}He/^{4}He$ appear to be high.

Another proposed explanation of the 'primordial' helium component is that it comes from the core. However, the only measurements of noble gas partitioning between silicate melt and iron melt under pressures up to 100 kilobars indicate that the partition coefficients are much less than unity and that they decrease systematically with increasing pressure (Matsuda et al., 1993). These results suggest that the Earth's core contains only negligible amounts of noble gases if core separation took place under equilibrium conditions at high pressure, although the uncertainties are too great to rule out the core as a potential mantle ^{3}He source (Porcelli and Halliday, 2001).

Finally, it has been proposed that a large amount of ^{3}He has been delivered to the Earth's surface by cosmic dust and then recycled to the mantle (Allegre et al., 1993; Anderson, 1993). Several arguments have been made against this proposal (Farley and Neroda, 1998); in particular, the dust flux is several orders of magnitude too low to account for the ^{3}He flux out of the mantle, and ^{3}He is expected to be degassed from the dust particles during subduction.

7.12.1.6.1.2 Heat–helium imbalance

The present-day flux of ^{4}He out of the mid-ocean ridge system has been estimated by measuring the concentration of ^{4}He in ocean water. O'Nions and Oxburgh (1983) pointed out that this ^{4}He flux is only a small fraction of what would be expected from the concentration of ^{238}U, ^{235}U, and ^{232}Th in MORB and even lower than U and Th in BSE, which has led to the term 'heat–helium imbalance.' The 'traditional' explanation of this is that the radiogenic sources are trapped in the lower mantle, from which heat can escape by conduction across a thermal boundary layer (TBL), but ^{4}He cannot. It has been suggested that the concentration of ^{4}He might be misleading because this is influenced by short-timescale processes such as variations in eruption rate, whereas the concentration of U and Th is a long-timescale process. Possible solutions to this are discussed in **Sections 7.12.4.1.1** and **7.12.5.1**.

7.12.1.6.1.3 Argon outgassing

Argon has three stable isotopes, the two of most interest being 'primitive' ^{36}Ar and 'radiogenic' ^{40}Ar. There is virtually no primordial ^{40}Ar in the Earth (Ozima and Kudo, 1972), it being entirely produced by the decay of ^{40}K. As Ar does not escape from the atmosphere (unlike He) due to its large atomic mass, the amount of ^{40}Ar in Earth's atmosphere represents the total amount of argon degassed from the Earth's interior over the age of the Earth. This atmospheric ^{40}Ar budget, when combined with an estimate of the total ^{40}K content of the Earth, implies that approximately half of all ^{40}Ar produced within the Earth since its formation is retained within the solid Earth (Allegre et al., 1996; O'Nions and Tolstikhin, 1996), which gives a constraint on the fraction of the mantle that has been processed through mid-ocean ridge or other melting environments. There is some uncertainty in this figure due to uncertainty in the amount of ^{40}K in the mantle and a fairly small uncertainty in how much of the ^{40}Ar in continental crust has actually outgassed. The uncertainty in total amount of K arises because in the planetary formation process, K is moderately volatile, so the total amount of K is usually calculated from the total amount of (refractory) U assuming some a particular K/U ratio, estimates of which range from a widely accepted value of 12 700 (Jochum et al., 1983) to as low as 2800 (Stacey, 1992). The continental crust contains ~1/3–1/2 of the K budget (Rudnick and Fountain, 1995) and for the earlier mass balance calculation is usually assumed to be completely outgassed in the resulting ^{40}Ar, which may not be the case (see, e.g., discussion in Phipps Morgan, 1998). The linkage between K assimilation by the continental crust and ^{40}Ar crust and mantle outgassing can be used to constrain the relative amounts of K in the crust and Ar in the atmosphere (Coltice et al., 2000a). In any case, this '50% outgassing' constraint is commonly misinterpreted to mean that 50% of the mantle is undegassed – in fact, because Earth's ^{40}Ar was continuously produced over geologic time from a negligible initial concentration, the mantle could have almost entirely degassed early in its history and still match the Ar constraint. In other words, early or accretionary degassing events do not influence ^{40}Ar outgassing. Quantitatively, for a constant convective vigor, van Keken and Ballentine (1999) showed that outgassing 50% radiogenic ^{40}Ar implies outgassing 72% of primordial, nonradiogenic species. If the convective vigor were higher in the past, then the latter fraction can be much higher. Phipps Morgan (1998) showed that the outgassing constraint is perfectly compatible with whole-mantle convection, even if the processing rate was much higher in the past: for example, assuming that the present-day processing time is 9.6 Gy and was faster according to $(\text{heating rate})^2$ in the past, one obtains 93% outgassing of primitive ^{36}Ar but only 58% outgassing of radiogenic ^{40}Ar, or 98% primitive outgassed while retaining 31% of radiogenic Ar. A better expression of this constraint is "50% of the mantle has been undegassed *since the ^{40}Ar was produced in it*" (Phipps Morgan, 1998).

7.12.1.6.1.4 The argon paradox

A more problematic constraint arises from the concentration of Ar in the MORB source region, which is much lower than what would be expected if the ^{40}Ar remaining in the mantle were evenly distributed (Turner, 1989). This has traditionally been interpreted to mean that the additional, 'missing' ^{40}Ar must be 'hidden' in the lower mantle, which should have about 50 times the ^{40}Ar concentration of the upper mantle (Allegre et al., 1996). It has been argued that this difficulty could be resolved if the amount of ^{40}K in the mantle were lower than commonly estimated (Albarède, 1998; Davies, 1999; Lyubetskaya and Korenaga, 2007b). Another reconciliation is that this, as well as the heat–helium imbalance and the apparent need for a deep ^{3}He reservoir, could be reconciled if the estimate of ^{40}Ar in the shallow mantle, which comes from measured short-term ^{3}He fluxes into the oceans, were too low by a factor of 3.5 (Ballentine et al., 2002). It is important to note that the possibility of K in

the core does not help with this paradox, because it is the imbalance between mantle Ar and K that is the problem.

7.12.1.6.1.5 Neon, xenon, and krypton

Far less attention has been paid to isotopes of these elements, largely because of atmospheric contamination problems (Ballentine, 2002; Ballentine and Barford, 2000; Farley and Neroda, 1998), but some recent work has important implications for mantle processes. Neon has three isotopes (atomic weights 20, 21, and 22), of which ^{21}Ne is radiogenic and ^{20}Ne and ^{22}Ne are primordial. Measurements of ^{20}Ne/^{22}Ne in magmatic CO_2 well gases (Ballentine et al., 2005) provide the first unambiguous value for the convecting mantle noble gas abundance and isotopic composition and indicate that the upper mantle has a ^{20}Ne/^{22}Ne similar to that of solar wind-irradiated meteorites (~12.5, so-called Neon-B), which is significantly lower than ^{20}Ne/^{22}Ne of the solar nebula (13.6–13.8). This Neon-B value is reproduced in samples from Iceland and Hawaii (Trieloff et al., 2000) and Réunion (Hopp and Trieloff, 2005). Some studies (e.g., Dixon et al., 2000; Honda et al., 1999; Moreira et al., 2001) have argued for a solar component based on extrapolating plots of ^{20}Ne/^{22}Ne versus ^{21}Ne/^{22}Ne, but their measured ^{20}Ne/^{22}Ne values are not higher than the Neon-B value. However, some samples from Iceland (Harrison et al., 1999) and the Kola Peninsula (Yokochi and Marty, 2004) appear to have ^{20}Ne/^{22}Ne > 12.5, which if verified would represent a minor component.

These measurements imply that the convecting mantle is dominated by neon gained by solar wind irradiation of accreting material (Neon-B) but that another 'solar' component might exist, sometimes sampled by plumes, associated with early accreted material with a solar nebula composition. The latter might be in the deep mantle or core (Ballentine et al., 2005). Alteration by recycled material appears to have played a minimal role. While measured Xe and Kr ratios generally appear air-like (Ballentine, 2002), evidence for 'solar' Xe has also been detected in CO_2 well gases by Caffee et al. (1999). A recent analysis suggests that the mantle content of 'heavy' noble gases (i.e., Xe, Kr, and Ar) is dominated by ocean water subduction (Holland and Ballentine, 2006), consistent with the finding of Sarda et al. (1999) that a significant amount of mantle Ar is recycled. Subduction of oceanic lithosphere plays a strong role in generating heterogeneity in many isotopes, as discussed in the succeeding text.

7.12.1.6.2 The end-member 'zoo'

Analysis of the varying trace-element isotopic compositions of OIB and MORB, particularly those of Sr, Nd, Hf, and Pb, suggests that they are the result of mixing between several 'end-members,' which are typically referred to as DMM, HIMU, EM1, and EM2 (Carlson, 1994; Hofmann, 2003; White, 1985, 2010; Zindler and Hart, 1986). Additionally, considering He data leads to a further component, commonly referred to as FOZO. The common interpretations of these will be discussed later. Another approach to understanding heterogeneity and its origins has been to identify the key directions (rather than end-members) in isotopic space using a principal component analysis or independent component analysis, which suggest that the observed variation is the result of only a few processes: according to Stracke (2012), 94.3% of the isotopic variation in oceanic basalts can be fit by two principal components corresponding to the DMM–PREMA trend and Pitcairn-like enriched mantle (EM), and adding Samoa-like EM and HIMU brings the fit to 99.2%; the independent component analysis of Iwamori et al. (2010) finds that 99.4% of the variance can be explained by only 3 independent components, which may be caused by oceanic melting followed by crustal recycling and segregation, variations in the amount of water, and continental crust.

7.12.1.6.2.1 Depleted MORB mantle

This end-member represents the most depleted MORB. Most MORBs are not as depleted as DMM, but plot on the DMM–FOZO trend.

7.12.1.6.2.2 High μ

μ is the ratio ^{238}U/^{204}Pb. As U and Th decay into ^{207}Pb and ^{206}Pb with time, high μ leads to a high ratio of radiogenic to primordial lead, that is, ^{206}Pb/^{204}Pb – a 'radiogenic' signature. This component is typically interpreted to be recycled oceanic crust (Allegre and Turcotte, 1986; Chase, 1981; Hofmann, 1997; Hofmann and White, 1982). However, as U and Pb are approximately equally compatible (or U is slightly more incompatible) when making oceanic crust, some additional mechanisms must be invoked to fractionate U and Pb to make the HIMU signature. The invoked mechanisms are hydrothermal removal of Pb from the subducted slab into subduction-related magmas (Hofmann, 1988) and enrichment of the oceanic crust in continental U carried by eroded sediments.

7.12.1.6.2.3 Enriched Mantle 1 and 2

These signatures are observed in OIBs, but their origin is still under discussion. Leading explanations are Carlson (1994) and Hofmann (2003): EM1 may originate from either recycling of delaminated subcontinental lithosphere or recycling of subducted ancient pelagic sediment (e.g., Rapp et al., 2008), while EM2 may be recycled oceanic crust with a small amount of sediment, melt-impregnated oceanic lithosphere, or sediment contamination of plume-derived magmas as they pass through the crust.

7.12.1.6.2.4 Focus zone

This corresponds to the high ^{3}He/^{4}He component and in other isotopes has a composition on or close to the DMM–HIMU trend. Samples from many OIB associations appear to form linear trends that radiate from focus zone (FOZO) to various enriched compositions (Hart et al., 1992; Hilton et al., 1999). FOZO is also referred to as PREMA (prevalent mantle) (Stracke, 2012). Components that are very similar to FOZO are C (Hanan and Graham, 1996), PHEM (primitive helium mantle) (Farley et al., 1992), PREMA (Zindler and Hart, 1986), and HRDM (^{3}He recharged depleted mantle) (Stuart et al., 2003). A detailed discussion of the similarities and differences between these (as well as between FOZO and HIMU) is given by Stracke et al. (2005) and Stracke (2012), who argued that FOZO/PREMA is a ubiquitous component that is most likely related to subducted MORB/eclogite, something that is also concluded in the independent component analysis of isotopes in OIBs and MORBs by Iwamori and Albarede (2008).

7.12.1.6.2.5 Isotopic 'Ages'

As discussed earlier, isotope ratios evolve along separate paths in different materials due to fractionation of radiogenic parent and daughter elements during melting and subsequent radiogenic ingrowth of daughter isotopes. The resulting slopes on isotope diagrams indicate an age of fractionation and are widely used to date geologic processes. Even though the mantle has undergone continuous differentiation rather than a single differentiation event, plots of $^{207}Pb/^{204}Pb$–$^{206}Pb/^{204}Pb$ from MORB and OIB samples display a coherent line, or 'pseudoisochron,' from which an 'effective age' can be calculated. The Pb system has the advantage that Pb isotopes do not fractionate from each other on melting, which is a disadvantage with, for example, the Sm–Nd system. Pb–Pb plots display a trend from DMM to the highly radiogenic HIMU end-member. The Pb–Pb pseudo-isochron ages for various MORB and OIB groups correspond to effective ages of 1.5–2 Gy (Hofmann, 1997), although the exact meaning of these ages is not straightforwardly interpretable (Albarède, 2001; Hofmann, 1997). A recent mathematical analysis reveals that for ongoing differentiation, the Pb–Pb 'pseudo-isochron' age is generally much larger than the average age of crustal differentiation (Rudge, 2006).

7.12.1.6.2.6 Lead paradoxes

Another constraint is that all oceanic basalts and most MORB plot to the high $^{206}Pb/^{204}Pb$ side of the geochron, implying that the mantle has experienced a net increase in μ and implying that there is a hidden reservoir of unradiogenic (low μ) material. This problem is still unresolved and is discussed in more detail in Hofmann (2003) and Murphy et al. (2003). A second lead paradox, often called the kappa conundrum, is related to the discrepancy between measured $^{208}Pb/^{206}Pb$ ratios and the ratio of their parents, $\kappa = {}^{232}Th/^{238}U$. A layered-mantle 'steady-state' solution to this was proposed by Galer and O'Nions (1986), with a more recent layered solution proposed by Turcotte et al. (2001). Other recent work has focused on reconciling it with whole-mantle convection; a reasonable resolution is starting to recycle uranium into the mantle ~2.5 billion years ago (Elliott et al., 1999; Kramers and Tolstikhin, 1997).

7.12.1.6.2.7 Other isotopes

Several other isotopes give clues about long-term storage of oceanic crust and the origin of isotopic heterogeneity; some examples are given here. Cabral et al. (2013) used sulfur isotopes in plume lavas to infer subducted crust older than 2.45 billion years. Chauvel et al. (2008) used Hf–Nd array to infer that the source of OIBs must include subducted oceanic sediments as well as subducted MORB. Cottrell and Kelley (2013) examined the oxidation state of MORB, finding fine-scale variations in which the EM is more reduced than depleted mantle, which may be related to the action of carbon producing enriched, reduced magmas.

7.12.1.6.3 Nd isotopes and early differentiation and mixing processes

The Sm–Nd system has two decays that are of geochemical usefulness: ^{147}Sm decays to ^{143}Nd with a half-life of 103 Ga, and ^{146}Sm decays to ^{142}Nd with a half-life of 106 Ma. The latter system, in particular measurements of ^{142}Nd anomalies in Archean rocks, has in recent years given us great new insights into processes happening in the first few 100 Ma, as reviewed in Caro (2011).

Firstly, the presence of ^{142}Nd anomalies in some cratonic rocks indicates an early formation (4.39–4.42 Ga) of primordial crust. Several lines of evidence suggest that this was related to crystallization of a global magma ocean, nonfractional in the lower mantle but fractional in the upper mantle (see **Section 7.12.1.5.1**). This Hadean reservoir remixed on a timescale of >1 Ga (Debaille et al., 2013), leading to homogeneous ^{142}Nd ratios in the modern mantle.

Secondly, the ^{142}Nd composition of the measurable Earth is different from that of chondrites, for which two explanations have been proposed:

(i) The early formation of an EER that has never remixed (Boyet and Carlson, 2006; Carlson and Boyet, 2008) and is probably located at the base of the mantle. This could also hold other 'missing' trace elements discussed earlier. A problem with this is that it would be expected to be sampled in some mantle plumes, but there is no clear evidence for this.

(ii) A nonchondritic Earth. A nonchondritic Earth might result from impact erosion of early crust that was enriched in incompatible elements (O'Neill and Palme, 2008); the isotopic similarity of the Earth and Moon suggests that this would have occurred before the Moon-forming impact (Caro, 2011). Campbell and O'Neill (2012) pointed out that a nonchondritic Earth also helps resolve other geochemical paradoxes, including those related to He and Ar (**Section 7.12.1.6.1**).

More constraints are needed to distinguish which of the previously mentioned possibilities is correct. While Rizo et al. (2012) found geochemical evidence for a Hadean enriched reservoir in 3.7 Ga old rocks, this could be caused by proto-crust rather than EER. Evidence for a very old primitive reservoir with high $^{3}He/^{4}He$ and a Pb isotopic age of about 4.5 Ga was found in relatively recent rocks by Jackson et al. (2010), but these also do not distinguish the two possibilities: from ratios of Nd and other isotopes, the reservoir could be either the residual reservoir from EER formation or the original nonchondritic Earth.

7.12.1.7 Length Scales: Geochemical and Seismological

Of primary interest for understanding mantle mixing processes and the preservation of chemical end-members is the length-scale distribution of chemical heterogeneity in the mantle, so this section reviews constraints from geochemical and seismological observations. In general, it is observed that heterogeneity exists at all scales, but quantitative constraints on the spectrum are minimal.

7.12.1.7.1 Geochemical scales

Chemical heterogeneity is observed at all scales (Carlson, 1987, 1994; Meibom and Anderson, 2004), even within the relatively uniform MORB source. For example, at the smallest

scale (i.e., melt inclusions in olivine phenocrysts), Pb isotope ratios in OIB span a large fraction of the global range (Hauri, 2002; Saal et al., 1998; Sobolev, 1996; Sobolev et al., 2000; Sobolev and Shimizu, 1993), while Nd ratios in the Ronda peridotite massif in Spain exceed the global OIB Nd ratio variation (Reisberg and Zindler, 1986). According to Sobolev et al. (2000), extreme heterogeneity of trace-element patterns in olivine phenocrysts at Hawaii indicates that the isotopic signature is maintained up to crustal magma chambers, limiting the scale of heterogeneities in the source region to a few kilometers or less. At the centimeter to meter scale, mantle rocks outcropping in peridotite massifs have been observed to contain centimeter to decimeter thick isotopically enriched pyroxenite veins embedded within a mostly depleted peridotite matrix (Pearson et al., 1993; Polve and Allegre, 1980; Reisberg and Zindler, 1986; Suen and Frey, 1987). Such variations have inspired the proposal of a 'marble cake' or 'plum pudding' style of chemical variation in the mantle, in which small volumes of enriched material reside within a matrix of depleted MORB mantle (DMM) (Allegre and Turcotte, 1986), a concept that has been the topic of much recent investigation as discussed in later sections. At the longest scale, there is also abundant evidence for compositional heterogeneity at scales of >1000 km in the form of the Dupal anomaly (Castillo, 1988; Dupre and Allegre, 1983; Hart, 1984; Klein et al., 1988; Schilling, 1973), a $\sim 10^8\,km^2$ region centered in the Indian Ocean that stretches in an almost continuous belt around the Southern Hemisphere between the equator and 60° S and is characterized by its anomalous Sr and Pb ratios.

Limited information exists about the spectrum of geochemical heterogeneity from sampling variations in isotope ratios along mid-ocean ridges. An early compilation of $^{87}Sr/^{86}Sr$ data from MORB and OIB (Gurnis, 1986b) seemed consistent with a flat or 'white' spectrum, but the wealth of data that have been gathered since then reveal more complex patterns. The spectrum of $^{3}He/^{4}He$ for 5800 km of the Southeast Indian Ridge displays periodicities of 150 and 400 km (Graham et al., 2001), which was interpreted to be indicative of upper mantle convection. Using new high-resolution MC-ICPMS data, Agranier et al. (2005) plotted the spectra of various isotope ratios along the Mid-Atlantic Ridge, finding two distinct signals. The first is a 6–10° length scale signal visible in Pb, Sr, and He ratios and the principal component 1, which is associated with the Iceland hot spot, while the second is a power-law spectrum with a slope of -1 visible in Nd, Hf ratios and principle component 2, which is perhaps indicative of dynamics stretching processes discussed later. Graham et al. (2006) found a bimodal distribution of hafnium isotope ratios along the Southeast Indian Ridge, which appears to indicate compositional striations with an average thickness of ~40 km in the upper mantle.

7.12.1.7.2 Seismological scales

The long-wavelength spectrum of heterogeneity is revealed by global seismic tomographic models, but much of the heterogeneity at this wavelength is thermal. There have, however, been attempts to separate thermal and chemical components to the seismic velocity variations (e.g., Forte and Mitrovica, 2001; Ishii and Tromp, 1999; Mosca et al., 2012; Romanowicz, 2001; Simmons et al., 2010; Trampert et al., 2004). Particularly notable are those studies that use normal mode splitting data because these are sensitive to density variations, providing an additional constraint. In particular, at the longest wavelengths of 1000s km, Ishii and Tromp (1999), Trampert et al. (2004), and Mosca et al. (2012) found that a substantial fraction of the shear wave heterogeneity in the deepest mantle is due to compositional variations and that the two low-velocity 'megaplumes' seen in global tomographic models seem to have dense material at their bases. Short-wavelength heterogeneity is constrained by scattering: short-period precursors to PKP reveal the presence of small ~8 km, weak (1% rms.) velocity perturbations throughout the mantle (Hedlin et al., 1997). There are strong regional differences in scattering strength (Hedlin and Shearer, 2000), with strong scattering beneath South and Central America, eastern Europe, and Indonesia, and some correlation with large-scale anomalies revealed by seismic tomography including the African superplume and Tethys Trench. Other types of seismic observation find structure with a gradient in velocity that is too large to be explained by diffusive thermal structures, including a dipping low-velocity layer in the mid lower mantle (Kaneshima and Helffrich, 1999) and sharp sides to the deep mantle 'superplumes' beneath Africa and the Pacific (Ni et al., 2002). Kaneshima and Helffrich (2009) found scatters of size ~10 km in the lower mantle at depths 1100–1800 km beneath the Pacific subduction zones, consistent with folded MORB. In summary, while there is ample seismic evidence for chemical heterogeneity in the mantle, it is not sufficient at this point to constrain the spectrum over all wavelengths, and in particular not to the sub-kilometer wavelengths to which subducted MORB is expected to be stretched.

The spectrum of heterogeneity expected from stirring theory is discussed in **Section 7.12.2.7**.

7.12.1.8 Geophysical Constraints in Favor of Whole-Mantle Convection

A number of types of geophysical evidence argue that mantle convection is not completely layered at the 660 km discontinuity, which is in conflict with the earliest models based on mass balance of Ar and other trace elements discussed earlier.

Firstly, seismological observations have long favored penetration of slabs into the lower mantle. While arguments initially revolved around interpreting travel times (Creager and Jordan, 1984, 1986), subsequent three-dimensional regional or global seismic tomographic models (Fukao, 1992; Fukao et al., 1992; Grand, 1994; Masters et al., 2000; Van der Hilst et al., 1997) clearly show slabs penetrating immediately in some areas or penetrating after a period of stagnation in other areas, although it has also been argued that thermal coupling in a two-layered convection could produce similar features (e.g., Cizkova et al., 1999). Significant flow stratification should lead to a strong decorrelation in global tomographic models across 660 km depth, which is not observed (Jordan et al., 1993; Puster and Jordan, 1994, 1997; Puster et al., 1995). Subsequent tomographic models have also revealed some upwelling plumes that appear to be continuous across the whole depth of the mantle (Montelli et al., 2004), although there is an active debate on this topic. Recent regional tomography appears to show plumes extending to at least the top of the lower mantle beneath (Wolfe et al., 2011) and Iceland (Shen et al., 1998), and both tomography and

660 km boundary topography indicate a plume below Yellowstone (Schmandt et al., 2012; Smith et al., 2009).

Several types of geodynamic evidence also favor penetration of flow into the lower mantle. Firstly, if the 660 km discontinuity were a total barrier to convection, then there would be a strong TBL at the base of the upper mantle, and all of the heat coming from the lower mantle would be carried by strong upwellings in the upper mantle. As demonstrated by Davies (1988), such pervasive strong hot upwellings would violate constraints based on the topography and geoid of the ocean floors, which instead appear to be passively spreading. Identifiable hot-spot swells assumed to be caused by upwelling plumes only account for 6–10% of the global heat flow (Davies, 1988; Sleep, 1990). Secondly, modeling the global geoid and dynamic topography using driving forces derived from seismic tomography indicates a better fit to observations with whole-mantle flow than with layered flow (Hager et al., 1985; Hager and Richards, 1989; Ricard et al., 1988, 1989), although some successful fits are possible with layered models (Wen and Anderson, 1997). A third line of evidence comes from correlating global tomographic models with the spatial locations of subduction zones in the past 140–180 Ma (Kyvalova et al., 1995; Richards and Engebretson, 1992; Scrivner and Anderson, 1992): Past subduction appears to match well with lower mantle tomography, implying that the subducted slabs are in the lower mantle. Extending this approach, 3-D mantle structure models have been constructed by combining historical plate reconstructions with a global flow solver, and the results show a reasonable match to fast structures observed in global tomographic models (Bunge et al., 1998; Deparis et al., 1995; Megnin et al., 1997; van der Meer et al., 2009). Finally, numerical convection models indicate that the endothermic phase transition at 660 km depth is too weak to completely layer the flow at present-day convective vigor – only intermittent regional layering is plausible (Bunge et al., 1997; Christensen and Yuen, 1985; Machetel and Weber, 1991; Peltier and Solheim, 1992; Tackley et al., 1994), although this does not rule out chemical layering per se, as chemical layering with a large enough density contrasts is additive to the phase change buoyancy effect (Christensen and Yuen, 1984).

While complete layering at 660 km depth seems ruled out, there is evidence in favor of chemical layering in the lowest approximately few hundred km of the mantle, perhaps localized in certain regions. Firstly, the D″ region has long been found to be seismically highly heterogeneous (e.g., Lay and Garnero, 2004), which might require compositional variations to explain. Second, long-wavelength seismic tomographic models robustly find two large regions of low seismic velocity underneath Africa and the Pacific. While these might plausibly be smeared-out groups of narrow plumes (Schubert et al., 2004), seismic evidence indicates that these are not purely thermal in origin, including (i) the lack of correlation and anomalous scaling between *P*-wave and *S*-wave velocities in the deepest part of the mantle (Bolton and Masters, 2001; Ishii and Tromp, 2004; Kennett et al., 1998; Masters et al., 2000; Saltzer et al., 2004; Simmons et al., 2010), (ii) negative correlation between bulk sound velocity and shear wave velocity (Antolik et al., 2003; Kennett and Gorbatov, 2004; Su and Dziewonski, 1997), (iii) a positive density anomaly, according to some models that use normal mode splitting data (Ishii and Tromp, 1999; Mosca et al., 2012; Trampert et al., 2004), and (iv) sharp sides, which are difficult to maintain with a diffusive thermal field (Luo et al., 2001; Wen, 2001, 2002). Some thermochemical convection models discussed later produce structures that are consistent with these seismological observations.

It has also been argued that geophysical evidence favors a ~1300 km deep, highly undulating layer in the deep mantle (Kellogg et al., 1999; van der Hilst and Karason, 1999). The boundary of this should, however, produce strong observable signatures (heterogeneity and scattering) that have not presently been found (Castle and van der Hilst, 2003a, b; Tackley, 2002; Vidale et al., 2001). At the present time, the evidence favors a deeper layer that is probably not global (see also **Chapter 7.11**).

7.12.1.9 Overview of Geochemical Mantle Models

Before discussing the geophysical findings regarding thermochemical convection, it is useful to give an overview of the major models that have been proposed to explain geochemical observations and reconcile geochemical and geophysical observations. Some of these are discussed in more detail in the final section. Rather than discuss a large number of proposed models, some of which have similarities, it is useful to discuss the two main characteristics that differ among models and that tend to be the aspects that models focus on. **Figure 2** schematically represents the main models.

7.12.1.9.1 Vertical layering

7.12.1.9.1.1 Layered at 660 km

The traditional layered model is chemically and dynamically stratified at the 660 km discontinuity (e.g., slabs do not penetrate), with the upper mantle being a depleted, degassed region that is the complement of the continental crust, and the lower mantle being primitive – the repository for 50% of the BSE's argon, for the high $^{3}He/^{4}He$ component, and for other trace elements that are missing from the (MORB source + continental crust) according to mass balance calculations (Allegre et al., 1996). Variations on this model include intermittent breakdown of layering (Stein and Hofmann, 1994), the idea that this layering broke down in recent history (Allegre, 1997), and the model of Anderson (1995), in which the upper mantle is divided into an enriched shallow layer (the source of OIBs) and a depleted transition zone (the source of MORB). This model is contradicted by a number of geophysical observations, as discussed in **Section 7.12.1.8**.

7.12.1.9.1.2 Deep, highly undulating layer (hidden anomalous layer)

This model, proposed by Kellogg et al. (1999), is conceptually similar to the traditional layered model except that the layer boundary is moved downward to ~1300 km above the CMB, undulating by ~1000 km. Recycled components reside on top of the layer. The layer is 'hidden' because the seismic wave velocity anomalies caused by the different composition and higher temperature cancel out. While this proposal has been thought-provoking and has stimulated increasing progress toward findings a unified model, it should produce a seismic signature that is not found, as discussed in **Section 7.12.1.8**.

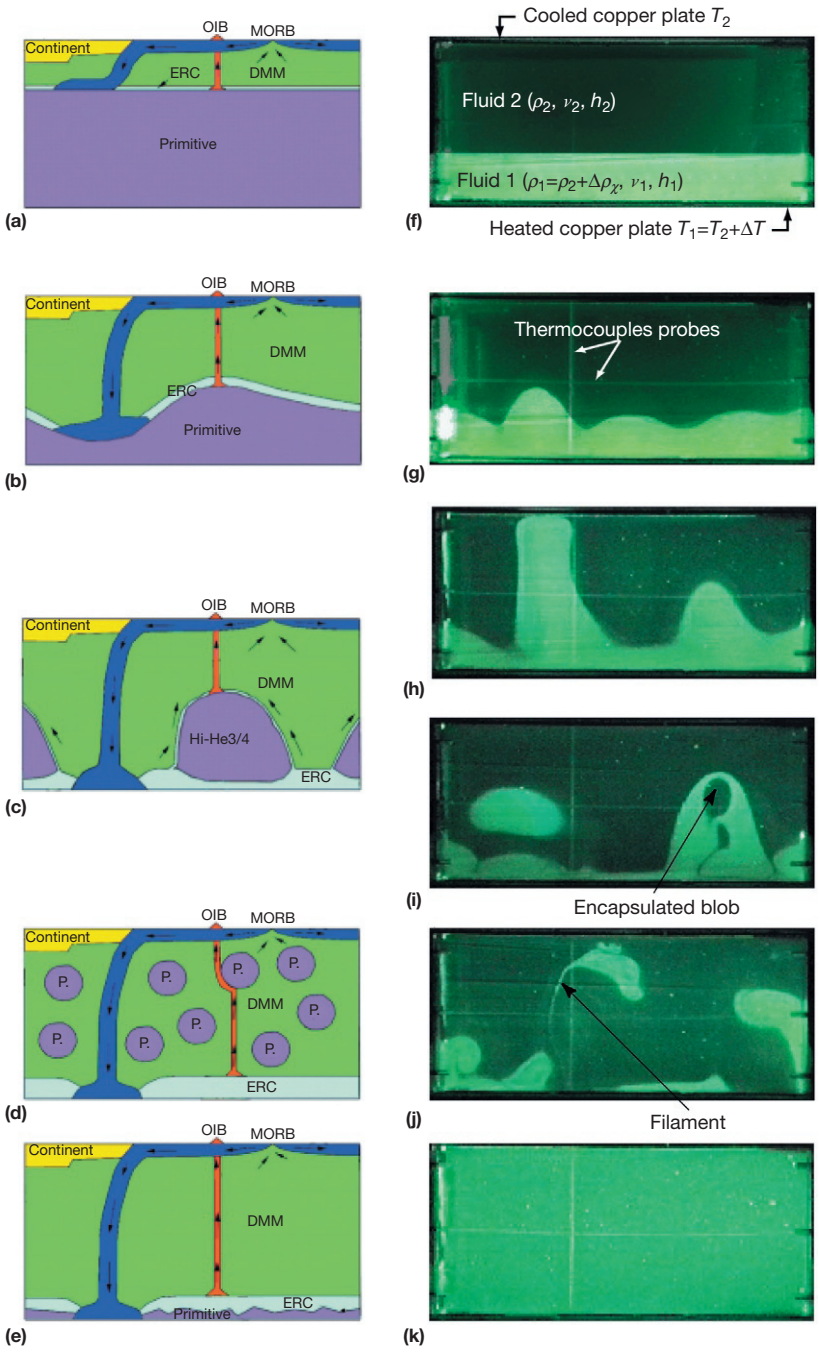


Figure 2 Various mantle models proposed to explain geochemical observations compared to results from a laboratory experiment at different times in its evolution, taken from Le Bars and Davaille (2004b). Cartoons in the left column are from Tackley (2000). (a) Layered at the 660 km discontinuity, (b) deep, high-topography hidden layer (Kellogg et al., 1999), (c) 'piles' (Tackley, 1998) or 'domes' (Davaille, 1999a), (d) primitive blobs (Becker et al., 1999; Davies, 1984; du Vignaux and Fleitout, 2001; Manga, 1996), (e) whole-mantle convection with thin dense layer at the base. (f)-(k) show the results of a laboratory experiment, which goes through different phases that display similar structure to the cartoons in the left column. Time increases from top (f) to bottom (k).

7.12.1.9.1.3 Thin layer (D″ region)

A approximately few hundred km thick layer above the CMB, perhaps highly undulating and/or localized into certain areas, and which could potentially contain primitive material, recycled material, or some mixture of the two, is a possibility. It has long been proposed that such a layer could contain recycled oceanic crust that provides the HIMU signature (Hofmann and White, 1982). Such oceanic crust contains a higher concentration of incompatible trace elements than primitive material so could also account for the 'missing' trace elements (about 1/3 of the total inventory of heat-producing elements according to Coltice and Ricard (1999)).

Whether such a layer could exist depends on the density contrast of subducted MORB at these high pressures and temperatures, as discussed in Section 7.12.3.3.2. It has been argued that this enriched layer may have formed shortly after the Earth's accretion process (Boyet and Carlson, 2005; Tolstikhin and Hofmann, 2005; Tolstikhin et al., 2006). Another proposal is that the layer preferentially contains oceanic crust containing oceanic plateaus, with normal oceanic crust not reaching this depth (Albarède and van der Hilst, 2002). Seismological observations support the existence of an intermittent layer that may extend several hundred km above the CMB, as discussed in Section 7.12.1.8.

7.12.1.9.2 Distributed heterogeneity

Instead of, or as well as, radial stratification, lateral compositional heterogeneity is thought to exist. An early justification for this was that mantle rocks outcropping in peridotite massifs have been observed to contain centimeter to decimeter thick isotopically enriched pyroxenite veins embedded within a mostly depleted peridotite matrix (Pearson et al., 1993; Polve and Allegre, 1980; Reisberg and Zindler, 1986; Suen and Frey, 1987). Such variations have inspired the proposal of a 'marble cake' or 'plum pudding' style of chemical variation in the mantle, in which small volumes of enriched material reside within a matrix of DMM (Allegre and Turcotte, 1986), a concept that has been the topic of much recent investigation as discussed in later sections, with several variations and refinements proposed. At the larger scale, it has been proposed that very large (~100 km scale) blobs may exist (Becker et al., 1999). Seismic scattering supports distributed heterogeneity, as discussed in Section 7.12.1.7. Dynamically, key issues are how rapidly distributed heterogeneity gets mixed (Section 7.12.2), which will be influenced by any composition dependence of viscosity (Section 7.12.2.3.5).

7.12.1.10 Mantle–Exosphere (Biosphere) Interactions: Water and Carbonate Cycling

Cycling of volatiles (particularly water and carbonate) between the interior and fluid envelope of a terrestrial planet could play a major role in the evolution of the fluid envelope and hence the long-term habitability of the surface environment. The mantle influences the fluid envelope by volatile outgassing associated with volcanism and by volatile recycling back into the interior. Indeed, a long-term cycle associated with carbonate recycling and reoutgassing may provide a critical feedback mechanism that maintains surface temperature in a habitable range (Kasting et al., 1993b; Sleep and Zahnle, 2001; Walker et al., 1981). Even on a plate tectonics timescale, changes in global subduction and spreading rates could influence climate (Lee et al., 2013). Volatile recycling also affects the redox state of the mantle, which, through outgassing, influences the redox state of the atmosphere (Delano, 2001; Kasting et al., 1993a; Lecuyer and Ricard, 1999; Trail et al., 2011; Zahnle et al., 2010), and may have been responsible for the rapid rise in atmospheric oxygen ~2 Gy ago (Kump et al., 2001). Crucial to recycling is the existence of plate tectonics, through which carbonate, water, and other volatiles are subducted into the mantle (e.g., Ruepke et al., 2004). Thus, the existence of plate tectonics on Earth may be important for long-term habitability.

Multiple feedbacks exist in the fluid envelope–mantle–core system. Both water and carbonate affect the physical properties of the mantle. Water strongly affects the viscosity of silicates (Hirth and Kohlstedt, 1996; although this has been disputed (Fei et al., 2013)), so water exchange between the mantle and exterior may have had important influences on the evolution of the coupled system. Furthermore, water greatly reduces the strength of the lithosphere through its effect on faults and the yield strength and is thus commonly thought to be necessary for plate tectonics to exist. Thus, the convective regime and heat flow out of the mantle are strongly affected by the fluid envelope through the influence of water on mantle viscosity and plate tectonics. Water also influences silicate melting, substantially reducing the solidus (see also Chapter 7.07).

Carbon is also important inside the mantle because it influences the physical properties, in particular having a major effect on melting and making water more mobile. In contrast to water, the amount of carbon in the Earth's interior is far greater than the amount in the exosphere. Detailed reviews of the origin and distribution of Earth's carbon and its influence on magma ocean and long-term processes can be found in Dasgupta (2013) and Dasgupta and Hirschmann (2010). Carbon's influence on silicate melting is greater than that of water, reducing the solidus by 100s K. Carbon can be present either as carbonated minerals or as diamond or graphite; indeed, Walter et al. (2011) found diamonds inferred to have formed in the lower mantle in subducted MORB. Carbonated melt could form in trace volumes and extract trace elements from great depth. Rohrbach and Schmidt (2011) found that carbonate-induced melting may occur in subducted lithosphere in the transition zone and lower mantle and freeze to form diamond. Later, diamonds could cause deep-seated melting at ~660 and ~250 km depth. Dasgupta et al. (2013) found that the depth of onset of melting under ridges is ~180 km for dry carbonated peridotite and 220–300 km for hydrated carbonated peridotite, although Stagno et al. (2013) obtained a shallower depth of 150 km due to a higher oxygen fugacity.

Modeling studies on the subduction of water and the influence of water on mantle thermal and geochemical evolution are discussed in Section 7.12.5.3. The influence of carbonate has so far been ignored in mantle modeling studies.

7.12.2 Mantle Stirring and Mixing of Passive Heterogeneities

7.12.2.1 Stirring, Stretching, Mixing, and Dispersal

It is important to distinguish between different concepts that are all sometimes loosely referred to as 'mixing.' Strictly speaking, 'mixing' involves homogenization of the chemical differences by a combination of stirring and chemical diffusion. Stirring involves stretching, which is the process by which an initial 'blob' of chemically distinct material is elongated into a thin tendril or lamella by strain caused by the convective flow. This greatly increases the surface area of the heterogeneity and decreases its width, facilitating chemical diffusion. Chemical diffusion of most elements is extremely slow, acting over length scales of centimeters to meters over geologic time; thus, it is necessary to stretch heterogeneities to this scale before they can be mixed. As will be discussed later, however,

observed MORB or OIB compositions probably involve averaging over a volume that is larger than this, so stretching to diffusion scales may not be necessary to eliminate the signature of an individual heterogeneity during melting. Furthermore, the presence of partial melt greatly increases the rate of chemical diffusion.

'Dispersal' refers to the dispersal of initially proximal heterogeneities evenly around the mantle. While this occurs partly by similar processes to stretching (i.e., the increasing separation of two initially very close points), effective dispersal involves material transport across convective 'cells.' It is possible in some flows for rapid stretching to occur within a cell, but for intercell transfer to be slow. Conversely, as will be discussed later, high-viscosity blobs undergo very slow stretching but may get rapidly dispersed around the domain.

Many of the concepts introduced in this section are illuminatingly illustrated in the review of van Keken et al. (2003). See also **Chapter 7.02**.

7.12.2.2 Stretching Theory: Laminar and Turbulent Regimes

7.12.2.2.1 Finite deformation tensor

The mathematical description of finite deformation was presented by McKenzie (1979) following Malvern (1969). We are interested in the deformation tensor relating the vector $\mathbf{y}'(t)$, which joins two particles in a fluid element at positions $x_1'(t)$ and $x_2'(t)$, to $\mathbf{y}$, the vector joining the same two particles at $t=0$:

$$\mathbf{y}'(t) = \mathbf{F}(t)\mathbf{y} \qquad [4]$$

The $\mathbf{F}$ matrix is initially the unit matrix and evolves according to

$$D_t F_{ij} = L_{ik} F_{kj} \qquad [5]$$

where D_t is the Lagrangian time derivative and $\mathbf{L}$ is the velocity gradient tensor:

$$L_{ij} \equiv \frac{\partial v_i}{\partial x_j} \qquad [6]$$

that can also be written in the material frame (Farnetani et al., 2002), which is a combination of the strain rate tensor (symmetrical part of $\mathbf{L}$) and the vorticity tensor (antisymmetric part of $\mathbf{L}$).

As the $\mathbf{F}$ tensor has nine components in three dimensions, or four components in two dimensions, it is desirable to extract a simpler quantity to describe the stretching. In two dimensions, a convenient scalar measure is the logarithm of the aspect ratio of the strain ellipse (McKenzie, 1979), that is,

$$f(t) = \log_{10}(a/b) \qquad [7]$$

where a and b are the major and minor axes of the ellipse, respectively, and a/b can be found from the second invariant of the $\mathbf{F}$ tensor, that is,

$$a/b = \gamma + \left(\gamma^2 - 1\right)^{1/2};\ \gamma = \frac{1}{2} F_{ij} F_{ij} \qquad [8]$$

Some authors prefer to measure the change in the semimajor or semiminor axis of the strain ellipse, which is the square root (or one over the square root) of (a/b). In two dimensions, Kellogg and Turcotte (1990) showed how a complete description of the total deformation of a heterogeneity can be obtained by tracking only two quantities for each tracer, the aspect ratio $(a/b \equiv e)$ and orientation θ of the strain ellipse. In three dimensions, the strain ellipse becomes an ellipsoid; a simplified method of tracking it is given in Subramanian et al. (2009).

7.12.2.2.2 Asymptotic stretching rate

Two types of shear flow, simple shear and pure shear, result in two different asymptotic stretching behaviors termed laminar and turbulent, respectively (see **Figure 3**). The theory is described by Olson et al. (1984a,b) based on previous theory by Batchelor (1952, 1959) and Corrsin (1961) and is briefly summarized here (see also Ricard, **Chapter 7.02**).

Laminar mixing (**Figure 3(a)**) occurs in a simple shear flow and involves a linear increase in the cumulative strain ε with time t:

$$\varepsilon = \dot{\varepsilon} t \qquad [9]$$

where $\dot{\varepsilon}$ is the average strain rate experienced by the heterogeneity. A linear decrease with time occurs for the component of the wave number vector that is perpendicular to the velocity:

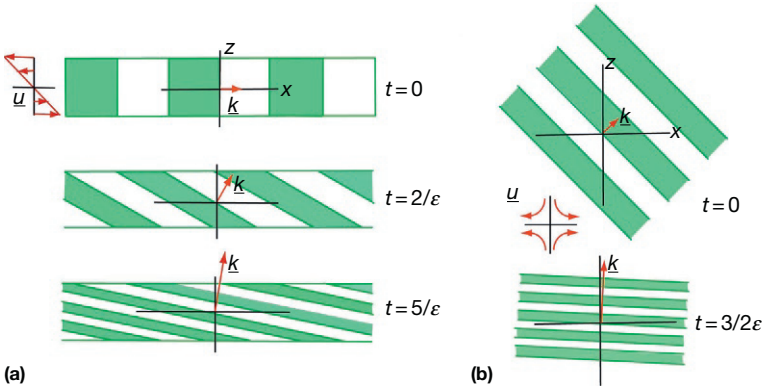


Figure 3 Diagrams illustrating the evolution of a single Fourier component of a passive anomaly for (a) uniform laminar shear flow and (b) stagnation point flow, based on Figures 4 and 5 of Olson et al. (1984b). The velocity field $\mathbf{u}$ is sketched for each case. In a uniform shear flow (a), the Fourier component is rotated into the direction of shear, while its wavelength decreases, such that the wave number magnitude $|\mathbf{k}|$ grows at a rate $\dot{\varepsilon}$, the Lagrangian strain rate. Near a stagnation point (b), the Fourier components also become rotated, this time into the direction of compression, but the wave number magnitude grows at a rate that is proportional to time, that is, $\dot{\varepsilon}t$.

$$k_x(t) = k_x(0);\ k_z(t) = k_z(0) + k(0)\dot{\varepsilon}t \qquad [10]$$

for a shear flow described as $\mathbf{u} = -\dot{\varepsilon}z\widehat{\mathbf{x}}$. This effectively describes a cascade of heterogeneity to shorter wavelengths, with 'blobs' being stretched into thin strips aligned with the direction of flow velocity. In the absence of diffusion, the Fourier amplitude of the heterogeneity does not change, it simply moves to higher k. For the case with diffusion, see the succeeding text.

Turbulent mixing is associated with pure shear (also known as normal strains (Olson et al., 1984a,b)), which occurs at stagnation points in convective flows (**Figure 3(b)**). In this case, the velocity is proportional to the distance from the stagnation point ($\mathbf{u} = -\dot{\varepsilon}z\widehat{\mathbf{x}} - \dot{\varepsilon}z\widehat{\mathbf{z}}$), so the cumulative strain increases exponentially with time:

$$\varepsilon = \exp(\dot{\varepsilon}t) \qquad [11]$$

with wave numbers evolving exponentially in both directions:

$$k_z(t) = k_z(0)\exp(\dot{\varepsilon}t);\ k_x(t) = k_x(0)\exp(-\dot{\varepsilon}t) \qquad [12]$$

As turbulent mixing is far more rapid than laminar mixing, it is important to establish which regime is most appropriate to characterize flows associated with mantle convection, so this has been a major focus of research. This issue has been investigated theoretically, for kinematically driven flows and for convectively driven flows. The general findings are that (i) the type of mixing can be quite sensitive to the exact flow, although greater time dependence results in a greater tendency toward turbulent mixing, and (ii) different 'blobs' may undergo substantially different rates of stretching in the same flow depending on their location in that flow. This is discussed in the next section.

Some semantic issues are worth noting: Firstly, the very high Prandtl number flow that the mantle undergoes should actually be termed laminar due to the unimportance of inertial terms; nevertheless, it can lead to the turbulent stretching regime. Secondly, turbulent mixing is often referred to as chaotic mixing, and laminar mixing is often referred to as regular mixing (see also **Chapter 7.02**).

7.12.2.2.3 Measures of stretching

Most measures of stretching are based on the elongation of an infinitesimal strain ellipse, that is, the ratio of the major axis to its original value (a/a_0), calculated using the equations in **Section 7.12.2.2.1** and tracked using infinitesimal tracer particles distributed throughout the domain. In two dimensions, the width of the strain ellipse is the inverse of this value. Values of (a/a_0) exhibit a considerable spread, so it is important to consider the range in values as well as the average. When the time-averaged stretching rate increases exponentially with time, the appropriate average to take is the median (Christensen, 1989a) or the geometric mean, which is related to the mean in logarithmic space. In this approach, the strain ellipse is always considered to be infinitesimal in size even if it reaches 10^{10} or larger, so these measures are not appropriate for estimating the separation of two initially closely spaced points, a quantity that cannot become larger than the size of the domain.

Other than the median or geometric mean stretching, it is useful to consider quantities that convey information about the distribution of strains. The statistical distribution of stretching can be plotted in various ways; for example, Kellogg and Turcotte (1990) plotted the probability distribution of strain (a/a_0 vs. number of tracers), the cumulative distribution (a/a_0 vs. fraction of particles exceeding (a/a_0)), and the time dependence of strain evolution (time vs. a/a_0 with different lines for the percentage of tracers exceeding a/a_0). An example of the latter quantity is the size reduction exceeded by 90% of tracers plotted by Christensen (1989a).

Often, the goal is to determine what fraction of heterogeneities has 'mixed,' in the sense of being stretched to less than the diffusive or sampling length scale. Kellogg and Turcotte (1990) plotted as a function of time the fraction of tracers that have been completely homogenized. Gurnis and Davies (1986b) were interested in heterogeneities that remained relatively intact, so they plotted as a function of time the fraction that has strain < 5.

7.12.2.2.3.1 Lyapunov exponents

The Lyapunov exponent is the exponential coefficient for the rate at which two initially close heterogeneities are separated. This can be written as

$$\lambda\left(P, \vec{\gamma}\right) = \lim_{\substack{t\to\infty \\ \gamma\to 0}} \left[\frac{1}{t}\ln\left(\frac{\gamma(t)}{\gamma}\right)\right] \qquad [13]$$

where γ is the length of separation vector $\vec{\gamma}$ located at position P and $\gamma(t)$ is its length after a time t. This is zero for laminar stretching and positive or negative for exponential stretching. The value depends on the direction of $\vec{\gamma}$ relative to the velocity gradient tensor, so λ has maximum and minimum values. Kellogg and Turcotte (1990) used the maximum Lyapunov exponent and the minimum Lyapunov exponent to quantify the stretching (Wolf, 1986), and these are straightforwardly related to $\overline{\alpha}_e$, the average strain rate measured from the numerical experiments, by

$$\lambda_{\max} = \frac{\overline{\alpha}_e}{\ln 2} \qquad [14]$$

where

$$\overline{\alpha}_e = \frac{1}{\tau}\ln\left(\frac{a}{a_0}\right) \text{ or } a = a_0\ \exp(\overline{\alpha}_e\tau) \qquad [15]$$

In numerical applications, it is often more convenient to calculate the Lyapunov exponent using finite rather than infinitesimal time difference t, or finite separation γ, for example, by considering the motion of nearby tracers. Lyapunov exponents calculated using these finite quantities are referred to as finite-time or finite size Lyapunov exponents, respectively, and have been used in the mantle dynamics literature (Farnetani et al., 2002; Farnetani and Samuel, 2003; Ferrachat and Ricard, 1998, 2001).

In general, for finite quantities, we can write using the previous equations that

$$\frac{\gamma(t)}{\gamma(0)} = \frac{|\mathbf{F}(t)\mathbf{y}|}{|\mathbf{y}|} = \left(\frac{\mathbf{y}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}(t)\mathbf{F}(t)\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\right)^{1/2} \qquad [16]$$

The maximum and minimum eigenvalues of the right Cauchy–Green strain tensor $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ can then be used to calculate the maximum and minimum finite-time Lyapunov exponents (e.g., Farnetani and Samuel, 2003):

$$\lambda_{max} = \frac{1}{t} \ln \sigma_{max};\ \lambda_{min} = \frac{1}{t} \ln \sigma_{min} \qquad [17]$$

In 2-D incompressible flow, $\lambda_{min} - \lambda_{max}$.

The hyperbolic persistence time method is a method to analyze stable or unstable material trajectories (unstable meaning that nearby particle trajectories separate from it) and was introduced to the mantle community by Farnetani and Samuel (2003). Readers are referred to this paper for more details.

7.12.2.3 Stretching: Numerical Results

7.12.2.3.1 General discussion

A convective flow includes both stagnation points (e.g., at the corners of a 'cell'), which will be associated with pure shear and exponential stretching, and laminar regions (away from corners), which will be associated with simple shear and linear stretching. It is therefore likely that a heterogeneity will experience linear stretching most of the time, with the occasional episode of exponential stretching. Which regime dominates the long-term behavior depends on the details of the flow. If the flow is steady state and two-dimensional, then most of the heterogeneities never pass close to a stagnation point, so the average stretching rate follows a linear stretching law (McKenzie, 1979; Olson et al., 1984b). If the flow is highly unsteady, then the stagnation points move around and new ones are generated, allowing essentially all heterogeneities to pass close to stagnation point at some point and giving on average an exponential stretching rate. The overall behavior is, however, highly sensitive to the details of the time dependence, and this has been studied by a variety of models ranging from purely kinematic flows to purely convective flows (e.g., Christensen, 1989a). As all heterogeneities are not stretched at the same rate, it has proven important to study the range of stretching rates as well as the average. Most of the research has been done in two dimensions, so three-dimensional studies are discussed in a later section.

7.12.2.3.2 Steady-state 2-D flows

Steady-state 2-D Bénard convection cells follow the laminar stretching law, whether kinematic (Olson et al., 1984b) or convective (McKenzie, 1979). The fit to theory is reasonable for both nondiffusive and diffusive heterogeneities (Olson et al., 1984b). It is instructive to understand the resulting stretching regime as a baseline for interpreting more complex flows. Heterogeneities are stretched into spirals of lamellae that are locally oriented subparallel to the streamlines, because different streamlines are rotating at a different rate. For example, near the top of the domain, the lamellae are locally subhorizontal. In a completely heated-from-below steady-state Bénard convection cell, the interior of the cell (far from the boundaries) undergoes almost rigid-body rotation with an associated low stretching rate, but near the boundaries, where heterogeneities pass close to stagnation points in the corners, the cumulative strain can increase substantially (e.g., the aspect ratio of the strain ellipse (a/b) by a factor of 1000) over one rotation (McKenzie, 1979). It is interesting to note that strain does not increase monotonically, but rather as large oscillations with a superimposed linear trend (McKenzie, 1979). The parts of these oscillations where the strain decreases have been termed 'unmixing.' The addition of internal heating makes the situation more complex as there is no uniformly rotating area in the middle, but again, the cumulative strain (a/b) undergoes considerable oscillation and near the edge of the 'cell' can reach a factor of 100 over one rotation (McKenzie, 1979). The highest strain rates (hence velocity gradients) occur near the downwelling. For finite-sized bodies (initially cylinders), Hoffman and McKenzie (1985) found that for pure basal heating, the amount of stretching was rapid for a blob initially situated near the edge of the flow, but with internal heating, the blob became deformed but not greatly stretched in one revolution.

Steady corner flows relevant to mid-ocean ridges and subduction zones have also been studied (McKenzie, 1979). The most severe strain ($a/b \sim 30$) was experienced in the corner above a subducting slab, with the principal axis of the strain ellipse aligned subparallel to the streamline. Again, strain does not necessarily increase monotonically with time.

7.12.2.3.3 Time-dependent 2-D flows

Various types of 2-D flow have been studied, including purely kinematic flows, purely convective flows, and convective flows with kinematic upper boundary conditions. This has led to the identification of three main stirring regimes (Christensen, 1989a): a regime in which all particles are undergoing exponential stirring, a 'slow' regime similar to the laminar regime, and a hybrid regime in which some particles/regions are undergoing turbulent stretching while some are undergoing slow laminar stretching. Fully time-dependent convection with no kinematic constraints has always been found to give turbulent mixing, whereas controlling the flow with a kinematic upper boundary condition or fully kinematic description may lead to the 'slow' or hybrid regimes, though some kinematic flows lead to the turbulent mixing regime (e.g., Kellogg and Turcotte, 1990). More discussion of the factors that lead to these regimes now follows.

7.12.2.3.3.1 Slow stirring regime

In this regime, most heterogeneities are stretched into long streaks, but clumps exist where the streaks get folded, leading to the simultaneous existence of tendrils with average stretching rate and blobs with very slow stretching (Gurnis, 1986c) that can survive for up to tens of transit times. This regime has been observed in kinematically prescribed flows consisting of two counterrotating cells with a boundary that moves smoothly and periodically with time (Christensen, 1989a; Gurnis, 1986c; Gurnis and Davies, 1986b). While visually the system may look reasonably well mixed, the average stretching rate is slow compared to the other regimes and lies somewhere between exponential and linear with time (Christensen, 1989a; Gurnis, 1986c). Gurnis and Davies (1986b) also obtained a regime visually similar to this with a combination of convection and kinematic plates, but the asymptotic stretching rate was not analyzed.

7.12.2.3.3.2 Exponential stirring regime

In this regime, all heterogeneities experience exponential stretching, although at different rates. There are no unmixed islands or persistent large blobs. All fully time-dependent purely convective (i.e., with no kinematic control) 2-D flows in the mantle convection literature that have been suitably analyzed

are in this regime, including those in Christensen (1989a) and Hoffman and McKenzie (1985). So are kinematic flows in which there is an oscillation between a single-cell and a three-cell structure, like those of Kellogg and Turcotte (1990) and simplified versions in Christensen (1989a). A characteristic that this kinematic flow shares with time-dependent convection is the creation of new cell boundaries, which can occur due to plume formation, sublithosphere convection or delamination, or new subduction zone formation. In some cases, it can also be obtained by two-cell kinematic flow with the boundary motion containing more than one frequency component (Christensen, 1989a).

It is important to note that the effective stretching rate that goes into the exponential stretching law (sometimes called the Lagrangian strain rate (Coltice, 2005)) is much lower than the mean strain rate of the flow – ~10–30 lower according to Christensen (1989a). In two-dimensional convective flows with internal heating up to $Ra = 10^{10}$ and with basal heating up to $Ra = 10^{9}$, Coltice (2005) found that the Lagrangian strain rate scales in proportion to the convective velocity, that is, $Ra^{1/2}$ and $Ra^{2/3}$, respectively, even though the higher Ra flows are more time-dependent, implying that the stretching rate is simply proportional to the rate at which material is moving, provided the flow is time-dependent. This finding also applied to cases with a 100- or 1000-fold viscosity jump at mid depth.

7.12.2.3.3.3 Hybrid stirring regime

In this regime, most of the flow is exponentially stirred but there exist unmixed regions or 'islands' in the flow; the histogram of strain distributions can exhibit a bimodal distribution with one peak at very low strains (Christensen, 1989a). In the experiments of Christensen (1989a), this was found in relatively few cases – in two of the kinematic flows with more than one driving frequency. Nevertheless, it is a mode that is commonly observed in mixing experiments performed outside the solid Earth science community (see Metcalfe et al., 1995 and references therein). The wide range of stretching rates observed in this mode and the 'slow stirring' mode earlier indicate the importance of considering the range rather than only the median or mean.

The statistical range of heterogeneity stretching in the turbulent mixing regime was studied by Kellogg and Turcotte (1990) using a chaotic kinematic flow with time dependence driven by the Lorenz equations. Specifically, the flow had two cells, and the boundary between moved smoothly but chaotically such that sometimes, there was only one cell. Into this, they injected a large number of particles with different starting positions, in order to obtain robust statistics regarding their stretching. Naturally, exponential (turbulent) stretching was obtained, but the amount of strain experienced by different tracers varied very widely, such that after, for example, five overturns, the semimajor axis of the strain ellipse divided by its initial radius (a/a_0) averaged 10^4 but varied between ~10^1 and ~10^7 (first to ninety-ninth percentile) or ~10^2–~10^6 (tenth to ninetieth percentile). In other words, the (logarithmic) variation in cumulative strain is as large as the average cumulative strain. It is thus not meaningful to talk about precise quantities such as 'mixing time' based on average stretching; everything must be qualified in terms of probability distributions. The effective strain rate for this stretching was found to be 13% of the maximum strain rate that occurred at the stagnation point in the corner of the cell.

7.12.2.3.4 3-D flows

It may be expected that mixing in three dimensions is quantitatively different from that in two dimensions, but so far, studies have been few and appear contradictory about the sign of the change. Two key concepts are relevant:

(i) *Toroidal motion*. 3-D flows can have toroidal motion – the component of motion associated with rotation about a vertical axis and strike-slip motion – whereas 2-D flows (in a vertical plane) cannot. In steady-state 2-D flows, the streamlines (particle paths) are always closed, but in 3-D, the presence of toroidal motion can cause chaotic, space-filling particle paths, even if the flow is stationary in time (Ottino, 1989). One way of viewing this, as discussed by Ferrachat and Ricard (1998), is that toroidal motion increases by one the number of degrees of freedom of the Hamiltonian, for which one degree of freedom gives regular mixing and two degrees of freedom give chaotic flow. For steady-state flows driven by an upper boundary condition mimicking a spreading center with a transform offset, this effect was demonstrated by Ferrachat and Ricard (1998). Areas of rapid, turbulent mixing, as well as islands of laminar mixing, were identified based on Poincaré sections and maps of finite-time Lyapunov exponents. Quantitatively, a 3-D flow with zero toroidal component was found to give similar Lyapunov exponent to 2-D flow, while the addition of a toroidal component increased the exponents by a factor of ~2. A large lower mantle viscosity tends to confine toroidal motion to the upper mantle (Ferrachat and Ricard, 1998; Gable et al., 1991).

(ii) *Plumes versus sheets*. Because 3-D upwellings (and downwellings if slabs are not included) can be columnar plumes rather than infinite sheets, they have a smaller effect on disrupting the flow pattern and causing intercell mixing. This was apparent in time-dependent, constant-viscosity convection at Rayleigh numbers up to 8×10^5 calculated by Schmalzl et al. (1996): after several overturns, tracers had dispersed efficiently within individual cells (as indicated by histograms of tracer vertical positions), but dispersal between cells was much slower in 3-D than in equivalent calculations in 2-D. However, a subsequent suite of calculations comparing 2-D and 3-D (poloidal only) flows up to much higher Rayleigh numbers found that 2-D and 3-D flows have approximately the same Lagrangian strain rate (Coltice and Schmalzl, 2006). It could be that Rayleigh number is a key factor that was unrealistically low in the earlier studies.

Similar numerical mixing experiments in 2-D and 3-D were performed by Davies (2002) and Huang and Davies (2007a), respectively. In these, passive heterogeneities introduced by subduction were tracked with a focus on sampling statistics. They found essentially no difference between 2-D and 3-D for the diagnostics studied. In summary, these results plus those of Coltice and Schmalzl (2006) imply that overall mixing efficiency in 3-D is likely to be similar to that in 2-D with the same parameters. Toroidal motion may influence this and needs to be studied more in the future.

7.12.2.3.4.1 Particle trajectories

It is worth considering the particle trajectories in 3-D flows. In steady-state 3-D convection with zero toroidal component, intercell transfer is zero because individual flow lines lie on a two-dimensional distorted (but closed) toroidal surface (Schmalzl et al., 1995). Individual streamlines are generally not closed in a small number of orbits around the torus, but rather, the surface is progressively filled by the trace of a single streamline. Extreme shear strains can occur but mixing is in 2-D. This is illustrated in **Figure 4**. With the addition of time dependence, boundary layer instabilities perturb these 'toroidal' structures, allowing for cross cell mixing (Schmalzl et al., 1996). Here, it is important to note that the 'toroidal paths' followed by particles in a steady-state flow are completely different to the 'toroidal motion' caused by surface rotation and strike-slip motion – indeed, toroidal motion breaks up the tori found with purely poloidal motion (Ferrachat and Ricard, 1998). The only relevant laboratory experiments were the steady-state experiments of Richter et al. (1982) in which for convective rolls, blobs were smeared out into spiral sheets or corkscrews in a superimposed lateral motion, with more complicated topologies for spoke pattern convection, but still within individual cells and not cross cell.

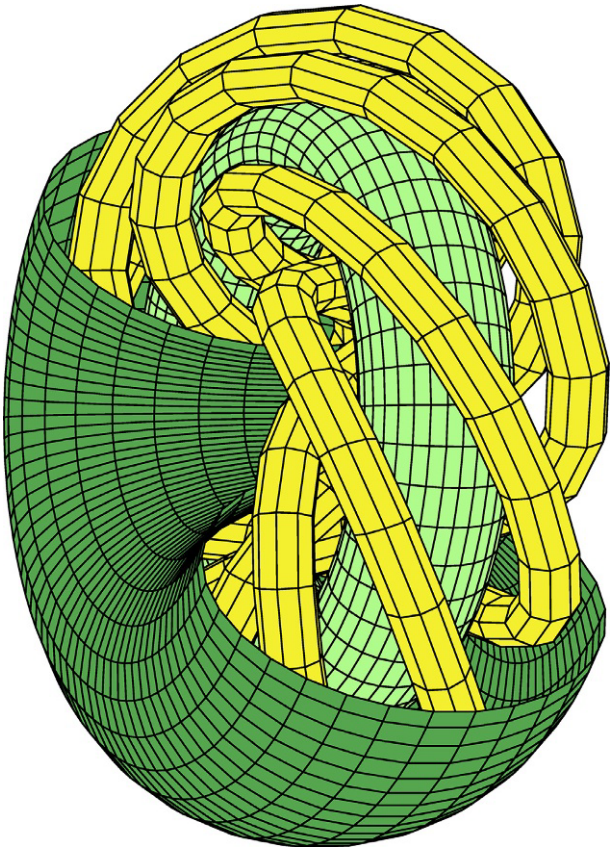


Figure 4 The topology of the pathlines for 3-D flow that is purely poloidal, extracted from a figure from Ferrachat and Ricard (1998). Individual tracer paths are restricted to tubes or tori, which they spiral round. Adding increasing amounts of toroidal motion breaks up the tubes and then the tori.

7.12.2.3.4.2 Spherical 3-D steady-state flows

The analysis of mixing in steady-state 3-D flows was extended to spherical geometry by van Keken and Zhong (1999), using a flow intended to be similar to the flow in the present-day mantle, that is, generated by buoyancy from a slab model combined with an approximate plate rheology. As with the simple flow of Ferrachat and Ricard (1998), only limited regions exhibited laminar mixing, with most other regions containing corkscrew-like particle tracks and probably chaotic mixing: The areas with highest stretching rates are those with high toroidal motion. Long-range dispersal was also studied, as discussed in a later section.

7.12.2.3.5 Effect of viscosity variations

7.12.2.3.5.1 Depth-dependent viscosity

For some time, it was thought that if the lower mantle has a substantially higher viscosity, then motion would be so sluggish that chemically distinct material could be retained for long periods of time (e.g., Davies, 1983). Numerical studies have, however, found that the effect of higher lower mantle viscosity on stirring and residence times is relatively small and that it cannot provide an explanation for long-lived heterogeneities per se. These numerical studies, which all study passive tracers, are discussed in this section.

Somewhat confusingly, the earliest study to address this issue (Gurnis and Davies, 1986a) did appear to show very slow stirring in a 100 times higher-viscosity lower mantle, even though the heterogeneities were passive, that is, had no density anomaly to keep them in the lower mantle. The circulation in these cases was however extremely sluggish and was arguably driven mainly by a prescribed platelike boundary condition rather than convection. (Although the authors were careful to make the prescribed surface velocities consistent with convection for constant-viscosity cases, this was not the case for depth-dependent viscosity cases (Christensen, 1989a, 1990).)

With a much higher convective vigor more relevant to the Earth and free-slip upper boundary, a lower mantle viscosity up to a factor of 100 higher than the upper mantle viscosity does not prevent large-scale mixing of the mantle (van Keken and Ballentine, 1998). The latter's calculations tracked helium degassing and ingrowth by radioactive decay, and it was found that the horizontally averaged %degassed and $^{3}He/^{4}He$ ratio were approximately constant with radius even for high-viscosity lower mantle cases, indicating good vertical mixing, although lateral heterogeneity was observed. Curiously, higher lower mantle viscosity seemed to cause higher outgassing rates – this was because the viscosity of the upper mantle was decreased in order to maintain the same surface heat flux in all cases, which led to higher velocities in the shallow mantle, hence more rapid outgassing. Subsequent models (van Keken and Ballentine, 1999) demonstrated that temperature-dependent viscosity and an endothermic phase change at 660 km depth also do not result in chemical stratification, even when the mantle starts out layered in trace-element content, although again, there was a significant amount of lateral heterogeneity (i.e., $^{3}He/^{4}He$ varied from 0 to a maximum value of 36–72).

Regarding stretching rate, a 100-fold viscosity jump was found by Coltice (2005) to make no difference to the average Lagrangian strain rate for internally heated convection, but for

basally heated convection, it decreased strain rate by a factor of ~4, probably because plumes from the lower boundary are affected by the viscosity jump, but instabilities from the upper boundary are not. In these calculations, the upper mantle viscosity and Rayleigh number were held constant.

Processing (residence) time was found to be insensitive to a viscosity jump, regardless of heating mode (Coltice, 2005). Consistent with random sampling and the He ratio results of van Keken and Ballentine (1998), Hunt and Kellogg (2001) found that a 100-fold viscosity jump at 660 km depth did not cause any age stratification. Seemingly inconsistent with Coltice (2005), Hunt and Kellogg (2001) found that the mean particle age increased by a factor of ~2, but this could be because they removed particles not only when they were sampled by melting but also when they became widely dispersed from other particles that were introduced at the same time, which means that the mean age cannot be interpreted as a simple residence time. The viscosity jump does, however, make a substantial difference to the ability of tracers to migrate laterally from their starting positions, because the flow is more steady, and this is something that should be studied further.

A nonconvective, steady-state flow, driven by top boundary forcing and assumed idealized density anomalies due to downwelling slabs, was studied by Ferrachat and Ricard (2001), who found that a 100-fold viscosity jump at mid depth increased the lower mantle mixing time (using a modified version of Olson et al. (1984a) definition based on even dispersion across boxes) by a factor of ~5 and that the finite-time Lyapunov exponents increased by a factor ~2–3. It is not clear how much such an idealized flow applies to the real Earth, however.

7.12.2.3.5.2 Composition-dependent viscosity

If a blob of compositionally distinct material has a higher viscosity than its surroundings, then the rate at which it gets stretched is considerably reduced. Simple theory predicts that in pure shear, the deformation rate is proportional to $1/(\lambda+1)$, where λ is the viscosity contrast (Manga, 1996; Spence et al., 1988). In a steady-state 2-D convective flow, Manga (1996) found that the deformation rate of extremely large (diameter ~1000 km) blobs was dramatically reduced with only a factor of ~10 viscosity increase, whereas if the viscosity of the blob is lower than its surroundings, then it rapidly gets stretched into tendrils. Manga (1996) also found that these large high-viscosity blobs tend to aggregate, raising the possibility that they could combine into even larger blobs.

For time-dependent convection (with a prescribed time-dependent top velocity condition), du Vignaux and Fleitout (2001) showed that stretching rate is more highly dependent on viscosity ratio than in simple shear and that the stretching is exponential with time, consistent with the turbulent mixing regime. Specifically, they found that

$$e_{\mathrm{m}} \equiv \frac{\mathrm{d}\ln(a^*/a_0)}{\mathrm{d}t} = \frac{C}{(1+\lambda)^2} \qquad [18]$$

where e_{m} is the mixing efficiency, a^* is the mean distance between two points initially at separation a_0, and C is a constant. The quantity a^* is commonly used in mixing studies as a measure of stretching (Ottino, 1989) and is related to, but not the same as, the semimajor axis of the strain ellipse a (du Vignaux and Fleitout, 2001). This equation is valid for viscosity contrast λ of approximately an order of magnitude or more. This scaling can be explained using an analytic theory of a blob experiencing repeated pure shear events in random directions (du Vignaux and Fleitout, 2001). While one might intuitively expect repeated shear in random directions to average out, or at least add like a random walk, the fact that when the major axis of the ellipse is at an angle like 45° to the pure shear direction, the stretching increases regardless of its orientation means that the overall effect is one of exponentially increasing strain.

Thus, differing viscosity of components can have a dramatic effect on the rate at which they get stretched and hence mixed. It is quite likely that chemically different components have different viscosities. Subducted basaltic crust in the upper mantle has a large proportion of garnet, which has a relatively high viscosity (Karato et al., 1995). The presence of water can reduce viscosity by up to approximately two orders of magnitude (Hirth and Kohlstedt, 1996), which will increase the viscosity of depleted residue and possibly decrease the viscosity of primitive material, which is rich in volatiles. In the lower mantle, magnesiowüstite likely has a lower viscosity than perovskite (Yamazaki and Karato, 2001), so assemblages that have different proportions of these minerals will have differing viscosity. Grain size also has a major influence on viscosity such that 'old' heterogeneities could have a higher viscosity (Solomatov and Reese, 2008).

7.12.2.3.5.3 Non-Newtonian viscosity

Mineral physics results suggest that a substantial part of the upper mantle is undergoing non-Newtonian dislocation creep (Karato and Wu, 1993), yet very few studies have addressed the influence of non-Newtonian power-law rheology on mixing. In a series of papers, Ten et al. (1996, 1997, 1998) compared the mixing properties of highly time-dependent Newtonian and non-Newtonian flows of similar vigor, in 2-D models that also had temperature- and pressure-dependent viscosity. Their basic finding is that mixing is less efficient in non-Newtonian flows, with unmixed 'islands' persisting for long time periods, in both plumes (Ten et al., 1996) and global convection (Ten et al., 1997, 1998). The local patterns of mixing were quite different between the different rheologies, with a greater richness in the scales of spatial heterogeneities in the non-Newtonian case (Ten et al., 1997). Non-Newtonian rheology causes chemical heterogeneity to persist for longer, implying that original chemical heterogeneities can remain unmixed for a long period and may be concentrated at certain depths (Ten et al., 1998). These results suggest that Newtonian models underestimate the mixing time in the upper mantle, something that warrants further study.

7.12.2.4 Dispersal

7.12.2.4.1 Measures of dispersal

7.12.2.4.1.1 Based on sampling cells

One class of measure for measuring the dispersal of initially close particles through the domain is based on dividing the domain into sampling cells and using statistical measures based on the number of particles in each cell. Initially, all the tracers will be in one cell, but when they are completely dispersed, the distribution of number of particles per cell will be indistinguishable from a random distribution.

Measures that have been used include the fraction of sampling bins with 0 tracers (Christensen, 1989a), or the fraction of sampling bins with greater than five times the average number of tracers (Christensen, 1989a) (both measures decrease with time), or the fraction of cells with >1 tracer, which increases with time (Schmalzl et al., 1996). Schmalzl et al. (1996) plotted statistics for each of several initial blocks and derived a 'rate of mixing' from the slope.

The statistics of the number of particles/cell can also be used to directly calculate a 'mixing time' (Olson et al., 1984a), which is the decay time for the reduction of the variance, as described in **Section 7.12.2.6.1**.

7.12.2.4.1.2 Based on separation

Given a distribution of particles, the *two-particle correlation function* $H(r)$ gives the fraction of pairs of particles with separation less than r (Schmalzl and Hansen, 1994). To calculate $H(r)$, all pairs of particles, that is, $N(N-1)/2$ combinations, must be considered. $H(0)$ is zero, and then $H(r)$ increases with r, reaching a maximum at either the size of the domain or the size of the tracer cloud (if smaller than the domain). The initial increase can generally be approximated as $H(r) \sim r^{\alpha}$, where α is the *correlation dimension*, and gives information about the dimensionality of the tracer cloud. In particular, $\alpha = 1$ corresponds to tracers spread into a line, $\alpha = 2$ to well-mixed 2-D convection, and $\alpha = 3$ to well-mixed 3-D convection. In spherical geometry, a random distribution of heterogeneities has a correlation dimension that depends on length scale, rising to a maximum of 2.8 at ~4000 km, falling monotonically to ~2 at ~7500 km, and then plunging to 0 by 10 000 km (Stegman et al., 2002).

The *root-mean square dispersion index* was introduced by Hoffman and McKenzie (1985) to quantify long-wavelength dispersion. It is defined as the square root of the second moment of the autocorrelation of the density distribution, the general form of which is

$$R_2 = \frac{\left(\iint r^2 \mathrm{d}V_1 \mathrm{d}V_2\right)^{1/2}}{\left(\iint \mathrm{d}V_1 \mathrm{d}V_2\right)^{1/2}} \qquad [19]$$

where r is the separation of two volume elements $\mathrm{d}V_1$ and $\mathrm{d}V_2$ of the same body. To study the one-dimensional horizontal dispersion of a body represented by N tracers in a 2-D domain, this is simplified to

$$R_2 H = \frac{1}{N}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij}^2\right)^{1/2} \qquad [20]$$

where x_{ij} is the horizontal separation of two particles i and j and this measure reaches an asymptotic value (for perfectly random mixing) of

$$R_2 H(\infty) = \left[1/6\left(N^2 - 1\right)\right]^{1/2} \qquad [21]$$

7.12.2.4.1.3 Other

Other quantities that have been used are the time spectrum of a tracer's vertical position, which gives the 'overturn time' of the convection (Schmalzl et al., 1996), and Poincaré sections (Ferrachat and Ricard, 1998; Schmalzl et al., 1995), that is, 2-D slices of a 3-D volume showing where the particle paths intersect it. These are useful for identifying whether motion is restricted to closed 2-D surfaces as in Schmalzl et al. (1995) or is chaotic (Ferrachat and Ricard, 1998).

7.12.2.4.2 Numerical results

Many discussions of 'mixing' actually refer to the dispersal of heterogeneities around the domain, rather than the process of stretching and diffusion previously discussed. This dispersal is often studied by placing an array of tracers in one part of the domain and studying how rapidly they spread out and become evenly distributed. A common observation in laboratory and numerical experiments (e.g., Richter et al., 1982; Schmalzl and Hansen, 1994) is that dispersal (and mixing) occurs rapidly inside a convective cell but much more slowly between one cell and another; that is, long-range lateral dispersal can be a relatively slow process. Another way of expressing this is that 'mixing' properties depend on spatial scale. The rate of cross cell dispersal depends on how time-dependent the flow is, for example, how time-dependent upwellings and downwellings at the cell boundaries are and whether and how often the cell pattern reorganizes. In the limit of steady-state flow, intercell dispersion is zero in 2-D flows and in 3-D flows with only poloidal motion, but nonzero if toroidal motion is present, as will be discussed later.

7.12.2.4.2.1 Intracell dispersal

The dispersal of tracers within a cell scales similarly to stretching. In the experiments of Christensen (1989a), using a variety of kinematic and convective flows as discussed earlier, stretching-related diagnostics and dispersion-related diagnostics were compared and found to obey similar behaviors, including the delineation of three regimes of mixing.

7.12.2.4.2.2 Intercell dispersal

For time-dependent 2-D flows, several numerical studies have studied the rate of lateral dispersion as a function of parameters. Convection that is completely internally heated is typically highly time-dependent with no long-lived cells, and Richter et al. (1982) found that at $Ra = 1.4 \times 10^6$, tracers were fairly rapidly (e.g., in 0.1 diffusion times) dispersed across a box of aspect ratio 8 in what visually resembles a diffusion process. Using an analytic time-dependent flow, they argued that rapid long-wavelength dispersal requires that the convective pattern change on a timescale comparable to the overturn time. Basally heated convection tends to have a steadier cellular pattern. Hoffman and McKenzie (1985) presented a series of calculations, also in an 8×1 box at $Ra = 1.4 \times 10^6$, with either 100% internal heating or 50% basal and 50% internal, and found that the flows with basal heating produced slower lateral dispersal, although it was still quite rapid. Completely heated-from-below flows were investigated by Schmalzl and Hansen (1994), who found that at range $Ra = 10^6$–10^8, heterogeneities within one cell are destroyed rapidly, while two adjacent cells can remain unmixed for substantially longer, although at $Ra = 3 \times 10^6$, particles were well dispersed by nondimensional time 0.02. They found the particle correlation function $H(r)$ and corresponding correlation dimension to be effective

measures of long-wavelength dispersion, with a value of 2 corresponding to a perfectly dispersed field. Higher Rayleigh number flows are more time-dependent such that the increase in the rate of laterally dispersion with *Ra* is higher than the increase in velocity with *Ra* (Schmalzl and Hansen, 1994).

7.12.2.4.2.3 3-D spherical results

Results regarding intercell dispersal in Cartesian 3-D flows were mentioned in **Section 7.12.2.3.4**. Long-range dispersal in steady-state 3-D spherical flows was studied by van Keken and Zhong (1999), using a flow intended to be similar to the flow in the present-day mantle, generated by buoyancy from a slab model combined with an approximate plate rheology. They found that particles could be efficiently transported far from their source: after 4 billion years of motion in this steady-state flow, pairs of particles initially 10 km apart were dispersed around the mantle with an almost flat probability distribution (i.e., number vs. separation). This suggests that dispersal in the actual mantle is relatively efficient. Stegman et al. (2002) extended such 3-D spherical modeling in order to determine whether dispersal is reduced by higher lower mantle viscosity (see **Figure 5**). Using the correlation dimension to measure dispersion, they found that in a model with 100 times higher lower mantle viscosity and driven by a slab model and present-day plate motions, the correlation dimension reached after a few billion years is only of order 1, which is normally indicative of stretching into a line. A representative structure can be seen in **Figure 5**. They also found that in a model with 100 times higher lower mantle viscosity and driven by a slab model and present-day plate motions, the upper mantle became mixed about 60% faster than the lower mantle, which mixing is determined by the correlation dimension. They conclude that this is insufficient differential mixing to explain the difference between the relatively high heterogeneity of OIBs and relatively low heterogeneity of MORBs, although it could be argued that the latter is related to the small-scale structures in the melting region, whereas the correlation dimension is quantifying variations at ~1000 km length scales.

7.12.2.4.3 Eddy diffusivity?

In the atmospheric community, long-wavelength lateral dispersal is commonly treated using an 'eddy diffusivity,' although it has been found to be dependent on the initial length scale (separation) being considered: larger separations lead to larger estimates of effective eddy diffusivity (Richardson, 1926; Richter et al., 1982). It is interesting to know whether this concept could be applied to the mantle. Olson et al. (1984b) studied whether an effective diffusion could be used to parameterize mixing within a convecting cell (an 'eddy') and concluded that it could not, because the effective diffusion coefficient would need to be larger for larger heterogeneity size. However, the two studies that have attempted to fit lateral dispersal across convective cells with an effective 'eddy' viscosity do appear to have had reasonable success.

For internally heated convection at $Ra = 1.4 \times 10^6$, Richter et al. (1982) used the distance-neighbor approach of Richardson (1926) to estimate an eddy diffusivity for the large-scale dispersal of heterogeneities at a scale length several times that of the depth and estimated a value of 30 times the thermal diffusivity, that is, ~0.003 $m^2 s^{-1}$, independent of initial separation. In similar experiments but with varying heating modes, Hoffman and McKenzie (1985) estimated a value of 25 times the thermal diffusivity for internally heated convection and 20 times the thermal diffusivity for 50% internal heated convection, by fitting the time evolution of the horizontal rms dispersion index with a diffusion curve, that is,

$$R_2 H = \sqrt{kt} \qquad [22]$$

The slight difference from the almost identical experiment of Richter et al. (1982) is probably due to difference in estimation technique and is not meaningful. They found that their

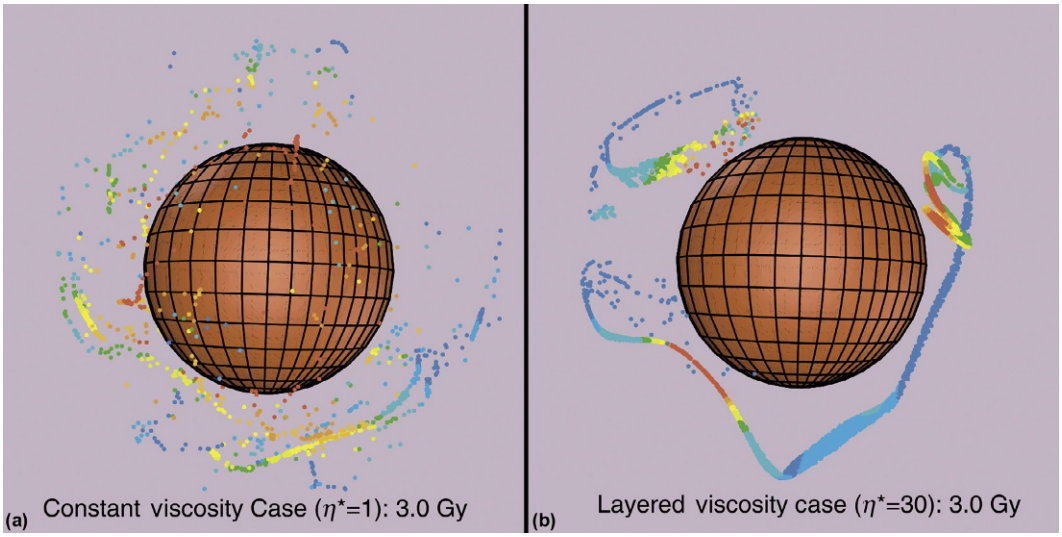


Figure 5 Dispersal of an initial 125 km radius cluster of tracers around the mantle after 3 Gy, from Stegman et al. (2002). Mantle flow is driven entirely by plate-like boundary conditions. The two panels show the effect of a 30* viscosity increase at the 660 km discontinuity. In the lower mantle, the heterogeneity is stretched into a ribbonlike structure, but this is more rapidly destroyed in the upper mantle or in an isoviscous mantle, where toroidal flow is more prominent.

numerical experiment agrees with this square root of time curve, so that eddy diffusivity is a good approximation for large scales of over 3 times the layer depth, and they got 20 times the thermal diffusion coefficient rather than 30 times. Based on these two studies, an eddy viscosity seems to offer a reasonable parameterization of lateral dispersal over distances greater than ~3 times the layer depth in a simple 2-D convective system, but the application to more complex situations and as a function of convective parameters remains to be determined.

7.12.2.5 Residence Time

The concept of residence time plays a central role in geochemical theory (e.g., Albarède, 1998) and also in numerical calculations of stirring and mixing processes. Residence time is defined as the length of time a heterogeneity (tracer) stays in the specified reservoir or other container. In geochemical box modeling, it may be applied to how long (on average) an isotope stays in a reservoir such as the continental crust and lower mantle. In numerical modeling of mantle evolution, it usually means the average length of time before a tracer is sampled by melting, from either the beginning of the calculation or, if tracers are being continually introduced at subduction zones, the average length of time before they are next sampled (melted). For whole-mantle convection in statistical equilibrium, this is equal to the processing time, that is, the time taken for one mantle mass to pass through the mid-ocean ridge processing zone, as discussed in **Section 7.12.1.5.3** where it was argued that this is around 6 Gy at the present processing rate but could have been much faster (i.e., lower) in the past.

From a chemical perspective, the mantle residence time for lithophile elements was calculated by Albarède (2005b) by dividing the amount of the element in a mean mantle model (i.e., BSE–continental crust) by the current rate at which the element is extracted into the oceanic crust. Residence times in the range 4–9 Gy were obtained, which bracket the geophysically constrained value of ~6 Gy discussed in **Section 7.12.1.5.3**. Again, it is likely that processing was substantially faster in the past.

In geochemical box modeling, it is commonly assumed that a reservoir is randomly sampled, that is, the probability of melting each atom or tracer in the reservoir is the same regardless of how long they have been in the reservoir, in which case the overall process is analogous to radioactive decay. Thus, if a reservoir starts off with N_0 atoms or tracers of a particular isotope, the rate of sampling is given by

$$\frac{dN}{dt} = \frac{N}{\tau} \quad [23]$$

where τ is the residence time. Hence, the number left after time t is given by

$$N = N_0 \exp(-t/\tau) \quad [24]$$

This residence time can be directly related to the mean age, where the age of a piece of mantle is defined as the time since it was last involved in a partial melting event or formation of the Earth. In equilibrium, of course, the mean age is equal to the processing time, but in the Earth, it is less than this because no piece of mantle can be older than the Earth and because the processing rate was likely faster (residence time was lower) in the past. For a constant processing rate, the mean age a is given by Huang and Davies (2007c):

$$a(t) = \tau(1 - \exp(t/\tau)) \quad [25]$$

For small times, this expression reduces to t and for long times, it asymptotes to τ, as expected. If the processing time is equal to 4.5 Gy, the mean age is thus ~2.8 Gy. If the processing rate were higher in the past according to (heating rate)^2, this reduces to 1.8 Gy, and if an increase in the depth of onset of melting is additionally taken into account, it reduces to 1.65 Gy (Xie and Tackley, 2004b).

The random sampling assumption is valid if tracers are effectively dispersed throughout the domain in question by time-dependent flows. This seems likely because (as discussed in **Section 7.12.2.3.3**) fully time-dependent free convection (i.e., without kinematic constraints) experiments have always displayed chaotic mixing, which effectively disperses tracers. This contrasts to steady-state 2-D flows, in which tracers are instead restricted to individual streamlines, only streamlines that intersect the sampling area are sampled with the result that amount of sampling increases linearly with time until all tracers on the those streamlines have been sampled (Coltice, 2005). Testing of this for mantle convection-like flows, as well as the influence of dense heterogeneities and viscosity variations, is now discussed.

The earliest study to check residence times was Gurnis and Davies (1986b), who investigated a low-Ra flow driven by internally heated convection at $Ra = 10^5$ and 'organized by' a kinematic platelike surface velocity condition. Tracers were introduced at the simulated moving trench, and it was found that the length of time tracers spend in the box before being sampled was easily predicted by the applied volumetric sampling rate and the rate of introduction of new tracers, consistent with the random sampling assumption. For more vigorous plate-organized flows in 2-D and 3-D, random sampling was verified by Davies (2002) and Huang and Davies (2007a,c), respectively. For free convection, Coltice (2005) confirmed that random sampling applies when the mixing is chaotic.

7.12.2.5.1 Effect of chemical density differences

While early models (Gurnis, 1986a) suggested that chemical density variations have little effect on sampling statistics, with a more Earth-like convective vigor, a significant effect is found (Christensen and Hofmann, 1994; Davies, 2002). Davies (2002) used roughly the correct convective vigor for Earth with chemical buoyancy ratios of up to 0.8 and found a significant difference for mid-ocean ridge sampling and for plume sampling. Mean residence times for MORB melting remained consistent with random sampling, with the mean age of tracers only slightly increased (from a scaled 2.42 to 2.52 Ga). However, some dense basaltic tracers settled into piles at the base, and their mean age increased by 32% from 2.57 to 3.38 Ga, with this higher age potentially being sampled by plume volcanism. If tracers are truly randomly sampled, then the distribution of ages should have a particular probability distribution that depends on the source function. Davies (2002) showed that tracers settling the base of the domain have a probability density function that is dramatically different from those in the

MORB sampling region, which implies that random sampling does not apply to all tracers even though the average residence time may not change much. Using a similar setup except in 3-D, Huang and Davies (2007a,b) verified that these conclusions also apply in 3-D. An appropriate description is probably that a different residence time applies to material that settles above the CMB; Christensen and Hofmann (1994) calculated such a quantity for calculations in which somewhat more basalt settling was obtained. They obtained residence times in the dense basal layer of between 0.36 and 1.62 Ga depending on parameters. For the whole system, they did not calculate the 'mean age' but rather the lead–lead pseudoisochron age; changing the buoyancy ratio from 1.0 to 2.25 while keeping other parameters constant changed the Pb–Pb age from 1.37 to 2.66 Ga. Thus, if density contrasts cause a layer to form, it is important to study residence times in the layer, and not just the average of the whole system.

7.12.2.5.2 Effect of viscosity variations and convective vigor

The effect of viscosity stratification and heating mode on residence time was determined by Coltice (2005). When the Ra is high enough for the flow to effectively disperse the tracers, the residence time scales with the theoretical prediction based on fraction of the mantle sampled per unit time, independent of viscosity contrast. This is consistent with earlier works that studied the effect of viscosity stratification on the depth distribution of mean age (Hunt and Kellogg, 2001; van Keken and Ballentine, 1998) – the effect is minimal. The 3-D spherical models of Huang and Davies (2007a,c) provided further evidence that viscosity stratification, temperature-dependent viscosity, convective vigor (Ra), and dimensionality do not have an effect on sampling statistics: the sampling is effectively random meaning that residence time depends only on the rate at which material is processed by partial melting.

7.12.2.6 Mixing Time

7.12.2.6.1 Background

Several definitions of 'mixing time' have been used, which, together with sensitivity of stirring rate to the exact type of assumed flow, has led to a wide variety of estimates for mixing time in the mantle. Some estimates have been based on stretching, while others have been based on dispersal. As discussed previously, 'mixing' (dispersal) across several convective cells can be far slower than 'mixing' in a single convective cell. In any case, a key quantity is the length scale of interest, which could be the diffusion length scale or sampling length scale.

The strictest definition of mixing time is the time required for the heterogeneities to become stretched thin enough to be homogenized by diffusion, which requires stretching down to submeter length scales for silicate rocks. It is questionable, however, whether such small-scale variations can be observed in geochemical observations (see later discussion), so an important concept is that of sampling volume (or, in 2-D, sampling area; Richter and Ribe, 1979), which can be reexpressed as a sampling length scale by taking the cubed (or square) root (Olson et al., 1984a,b). This sampling volume describes the region over which heterogeneities are homogenized by the sampling process, such as partial melting and extraction and mixing of magmas with each other and with the rock through which they are passing. While this has often been assumed to be approximately tens of kilometers, some of the geochemical observations discussed in Section 7.12.1.7 indicate that shorter-length-scale heterogeneity can be preserved in the resulting rocks. When calculating compositional heterogeneity in models, heterogeneity should be integrated (averaged) over the sample volume. In summary, the mixing time is the time required for heterogeneity to get stretched to the sampling size or the diffusion length scale, whichever is larger.

It has been usual to scale estimates of mixing time to either upper mantle convection or whole-mantle convection, which generally results in rather small estimated timescales for upper mantle mixing. These are probably irrelevant to the present-day mantle, as it is now clear that the upper mantle is not convecting independently as a simple heated-from-below Rayleigh–Bénard system, but instead, the convective regime is dominated by whole-mantle flow and the plates impose a long-wavelength organization to the upper mantle. It is possible that small-scale convection below the plates exists in the upper mantle (see Chapter 7.08), but this would be less vigorous than assumed in the 'upper mantle' mixing estimates and its effect on mixing has not yet been evaluated.

Some important statistical measures of heterogeneity are now introduced, following Olson et al. (1984a). The two-point correlation function is given by

$$R(\mathbf{r}, t) = \langle \theta(\mathbf{x}')\theta(\mathbf{x}' + \mathbf{r}) \rangle \quad [26]$$

where $\theta(\mathbf{x})$ is the compositional field, $\mathbf{x}$ is the position vector, $\mathbf{r}$ is the separation vector, and the angle brackets denote the volume average. Its Fourier transform, the energy spectral density (Olson et al., 1984a) in n spatial dimensions, is given by

$$E(\mathbf{k}, t) = \left(\frac{1}{\pi}\right)^n \int_0^\infty R(\mathbf{r}, t)\exp(-i\mathbf{k}\cdot\mathbf{r})\mathrm{d}\mathbf{r} \quad [27]$$

and the related variance of composition σ^2 is given by

$$\sigma^2 = R(0, t) = \int_0^\infty E(\mathbf{k}, \mathbf{t})\mathrm{d}\mathbf{k} \quad [28]$$

If heterogeneity is not created or destroyed but only changes wavelength, the variance remains constant with time (see Ricard, Chapter 7.02). However, if there is a sampling wavelength below which heterogeneity can no longer be detected, then it is appropriate to use the corresponding wave numbers as the upper limit for the integration, leading to the sampling variance

$$\overline{\sigma}^2 = \int_0^{k_s} E(\mathbf{k}, \mathbf{t})\mathrm{d}\mathbf{k} \quad [29]$$

which will decrease with time due the cascade of heterogeneity to shorter wavelengths if heterogeneity is not replenished at long wavelengths. This has led to one definition of the mixing time (Olson et al., 1984a):

$$\tau_{\mathrm{mix}} = \int_0^\infty \frac{\overline{\sigma}^2(t)}{\overline{\sigma}^2(t_0)}\mathrm{d}t \quad [30]$$

the physical meaning of which is the time needed to reduce the sample variance of the composition significantly. A modification of this should be made when calculating $\overline{\sigma}^2 = C_2(\kappa h_0/B^2)Ra^{1/3}$ numerically using a finite number of

sampling cells containing a finite number of tracer particles, as described by Ferrachat and Ricard (2001). In this situation, it does not tend asymptotically to zero, so the integral does not converge. Instead, the statistical limit is

$$\lim_{t\to\infty}\frac{\overline{\sigma}^2(t)}{\overline{\sigma}^2(t_0)}=\frac{s-1}{n} \quad [31]$$

where s is the number of sampling cells and n is the number of particles. Thus, Ferrachat and Ricard (2001) introduced a definition for mixing time that includes this and does converge:

$$\tau_{\text{mix}}=\int_0^\infty\left(\frac{\overline{\sigma}^2(t)}{\overline{\sigma}^2(t_0)}-\frac{s-1}{n}[1-\exp(-t/\tau_{\text{mix}})]\right)\mathrm{d}t \quad [32]$$

The mixing time is expected to depend on sampling length scale, as discussed earlier. To test this, the number of sample cells s can be varied, as was done by Ferrachat and Ricard (2001) and Olson et al. (1984a). Ferrachat and Ricard (2001) found that the mixing time increases regularly with 1/length scale, with a moderate increase in mixing time for decreasing cell size – approximately a factor of 4–5 for factor 32 decrease in length scale.

7.12.2.6.2 Laminar flows

A heterogeneity is stretched until it either is homogenized by diffusion or has a width of less than the sampling length scale. The balance between diffusion and sampling length scale in controlling the mixing time was analyzed by Olson et al. (1984b) for idealized flows and then compared to numerical results of a steady-state Rayleigh–Bénard convection cell in which the stretching rate is laminar with time. Substituting the spectral expansion of a heterogeneity into the previous equation led to the mixing time for nondiffusive laminar stirring being proportional to initial length scale/sampling length scale, that is,

$$\tau_{mix}\approx\frac{\pi}{\dot{\varepsilon}k_0\Delta z}=\frac{\lambda_0}{2\dot{\varepsilon}\Delta z} \quad [33]$$

where λ_0 is the initial wavelength of heterogeneity (if a localized heterogeneity, twice its size) and Δz is the sampling length scale. If, however, diffusion is rapid enough to homogenize the heterogeneity before it reaches the sampling length scale, a different expression is obtained:

$$\tau_{\text{mix}}\approx\left(\frac{\lambda_0^2}{4\kappa\dot{\varepsilon}^2}\right)^{1/3} \quad [34]$$

where κ is the diffusivity. There is a smooth transition between these types of behavior.

Olson et al. (1984b) tested these equations using a single Bénard convection cell and found that the results fitted reasonably well, for both nondiffusive and diffusive heterogeneities. Scaled to the Earth with an assumed $F=C/B^2$, an initial anomaly wavelength equal to the depth, and a sampling length scale of 5 km, this leads to mixing time of 2.3 Gy for upper mantle convection or 10 Gy for whole-mantle convection.

In the convective + kinematic internally heated ($Ra=10^5$) flows of Gurnis and Davies (1986b), long-lived clumps (folded strips of crust) that persisted over tens of transit times were observed, even though most tracers were well dispersed. Based on qualitative measures, the authors estimated the survival time of the clumps to be 2.4, or 1.4 Gy if faster processing in the past is taken into account.

7.12.2.6.3 Turbulent regime

For turbulent mixing, the equivalent result for nondiffusive flow is logarithmic and thus more rapid (Olson et al., 1984b):

$$\tau_{\text{mix}}\approx\frac{1}{\dot{\varepsilon}}\ln\left(\frac{2\lambda_0}{\Delta z}\right) \quad [35]$$

It is important to note that effective strain rate that goes into this equation can be more than an order of magnitude smaller than the average (e.g., root-mean-square) strain rate of the flow and depends on the details of the flow. For the convective flows investigated by Christensen (1989a) at Rayleigh numbers in the range 10^5–10^6, the effective strain rate was a factor of 10–30 less than the mean strain rate of the flow. The effective strain rate scales in proportion to convective velocity over a wide range of Rayleigh number for both basally heated and internally heated flows (Coltice, 2005).

As for mixing time, Christensen (1989a) assumed that the mantle is well mixed when >90% of the heterogeneities had been reduced to <1/50th of their original size and scaled his numerical results to the Earth using the transit time. This led to an estimate of 200–300 Ma for upper mantle convection or 1500–2000 My for whole-mantle convection with a reasonable viscosity increase with depth. There are two points to note: This mixing criterion involves stretching oceanic crust down to 100 m thickness, which is larger than what other estimates have assumed: if instead the mixing is taken to meter scale, this timescale estimate is approximately doubled to 3–4 Gy for whole-mantle convection. Secondly, there is a wide distribution in stretching rates, so that even when the 'mixed' criterion is met, there are still heterogeneities that are substantially larger and could influence isotopic signatures.

Hoffman and McKenzie (1985) also estimated mixing time from the scaling relationships that they found for small-scale stretching and large-scale dispersion. They estimated that 'any reasonable body' experiencing upper mantle convection will be thinned to 10^{-3} layer depths (700 m) in 400 My and 10^{-6} layer depths (0.7 m) in 700 My; for whole-mantle convection, these timescales are 1.0 Gy (for 3 km) and 1.5 Gy (for 3 m). However, lateral dispersion takes much longer: for half the circumference of the Earth and whole-mantle convection, the timescale is 4.6 Gy, similar to the age of the Earth.

The mixing time for subducted oceanic crust in a kinematic chaotic flow was calculated by Kellogg and Turcotte (1990). They assumed that a heterogeneity must be stretched thin enough for diffusive homogenization to take place based on the model of Kellogg and Turcotte (1986), which gives an equation for mixing time of

$$t_{\text{h}}=\frac{12h}{\pi u_{\text{sm}}\overline{\alpha}_{\text{e}}}\ln\left[4b_0\left(\frac{\pi u_{\text{sm}}\overline{\alpha}_{\text{e}}}{12Dh}\right)^{1/2}\right] \quad [36]$$

where D is the chemical diffusivity (assumed to be 10^{-19} m^2s^{-1} based on Hofmann and Hart (1978) and Sneeringer et al. (1984)), h is the depth of the convecting layer, b_0 is the initial heterogeneity size (assumed to be 6 km), u_{sm} is the average surface velocity (assumed to be 10 cm year^{-1}), and $\overline{\alpha}_{\text{e}}$ is the

effective strain rate from the numerical experiments. With these assumptions, t_h was calculated to be 240 and 960 million years for upper mantle convection and whole-mantle convection, respectively, which is the time for 50% of the heterogeneities to be homogenized by diffusion and corresponds to nine overturns. The diffusive homogenization length scale for this is <1 cm for both cases, and the amount of thinning is six to seven orders of magnitude. Due to idealizations in their model, however, these times must be interpreted as lower limits. One idealization is that this is not a real convective flow and it is not clear how representative the effective strain rate will be of the real mantle. For example, in the real mantle, the viscosity increases with depth, which will reduce the strain rate in the deep mantle and increase the 'overturn time,' a concept which itself is not well defined for flow in the actual mantle. An obvious quantitative issue is that the assumed average surface velocity of 10 cm year^{-1} is approximately three times larger than Earth's actual mean poloidal surface velocity, the poloidal component of which is around 3.5 cm year^{-1} (Lithgow-Bertelloni et al., 1993). If the times are recalculated with this value, they change to 660 and 2630 million years, respectively, the latter of which may well be consistent with geochemical evidence for survival of geochemical heterogeneity over 2 billion-year timescales.

In the actual mantle, heterogeneity is continuously introduced at the same time as 'old' heterogeneity is destroyed by stirring and diffusion, so Kellogg and Turcotte (1990) also considered the fraction of the mantle that is completely homogenized by the present day when oceanic crust is continuously introduced into the mantle over geologic time. For this, it is first necessary to assume an equation for crustal production versus time, and they assume that oceanic crustal production in the past scales with radiogenic heat production in the mantle with a $1/e$ decay time of 2.5 Gy (Turcotte and Schubert, 1982), that the present rate of production is 2.8 km^{2}year^{-1}, and that the processing depth under mid-ocean ridges is 60 km, assumptions that lead to half of the mantle being processed in 800 My for layered convection or 2200 My for whole-mantle convection. Combining this melting model with their mixing model, they derive an expression for $F(d)$, the fraction of particles narrower than d. Hence, setting d to the diffusion length scale of 10^{-7}, they obtain a curve for the fraction of particles that have been completely homogenized, in terms of the mixing time. For their estimated whole-mantle mixing time of 960 My, 67% of the mantle is perfectly mixed, but for the adjusted mixing time of 2.6 Gy as discussed earlier, this fraction is 28%, that is, over 70% of the present-day mantle is still in the form of strips of varying thickness. Thus, they predict substantial chemical heterogeneity even with the relatively rapid mixing in their experiments.

Kellogg and Stewart (1991) estimated the effective strain rate and mixing time for a chaotic flow using the same method, but this time, the flow is the nonlinear interaction of a single convection roll with a single spatial subharmonic, and the obtained mixing times are about a factor of 2 smaller.

7.12.2.6.4 Summary of different 'mixing time' estimates

Table 1 summarizes the various estimates of mixing time for whole-mantle convection, as discussed earlier. As far as the author is aware, these are all the published estimates given in dimensional time units, although it might be possible to obtain additional estimates by scaling nondimensional results from other studies. Estimates vary by an order of magnitude, from 1 to 10 Gy, because of different assumptions about the length scale to which heterogeneities should be stretched, the type of flow, and the flow velocity. Gurnis and Davies (1986b) were concerned with the survival of large clumps, so their estimate is not directly comparable to the other estimates. The estimate of Olson et al. (1984b) is large because they assumed laminar flow, whereas it is now established that time-dependent convective flows tend to give turbulent mixing, as occurred in the last three estimates. The estimate of Kellogg and Turcotte (1990) should be at least tripled because their assumed velocity was too large, as discussed in the preceding text. The estimate of Christensen (1989a) refers to stretching oceanic crust to ~100 m thickness and should be at least doubled for comparison with the Kellogg and Turcotte (1990) estimate. The estimate of Hoffman and McKenzie (1985) was made before it was established that the lower mantle is probably a factor of ~30 more viscous than the upper mantle (they assumed a factor of 3) so should be at least doubled when this is taken into account. When such scaling is done, estimates are in better agreement, in the 3–4 Gy range for stretching crust to the submeter length scale. If an equivalent table were constructed for upper mantle convection estimates, then the times would be much shorter, but as argued earlier, geophysical constraints indicate that the dominant mode of convection in the present Earth is whole mantle and the effect of possible 'small-scale' convection beneath the oceanic plates has not yet been evaluated.

7.12.2.7 Spectrum of Chemical Heterogeneity

As was discussed in Section 7.12.1.7, our knowledge of the spectrum of chemical heterogeneity in the mantle is far from perfect, but some useful information does exist. Geochemical

Table 1 Summary of mixing time estimates

Study	*τ_{mix} (Ga)*	*Length scale (inner)*	*Notes*	*Rescaled τ_{mix} (Gy)*
Gurnis and Davies (1986b)	2.4	Visible clumps	Survival time	
Olson et al. (1984b)	10	Diffusion	Laminar stretching	
Christensen (1989a)	1.5–2.0	1/50 (~100 m)	90% reduced to this size	3–4
Hoffman and McKenzie (1985)	1.0	3 km	only 3* viscosity increase with depth	>3
	1.5	3 m		
Kellogg and Turcotte (1990)	0.96	Diffusive	50% homogenized	~3

observations show the existence of heterogeneity at all length scales, from the scale of melt inclusions to thousands of kilometers, and quantitative information regarding the spectrum is becoming available. As also discussed earlier, seismological information points to the existence of scatterers throughout the lower mantle, with the amount of scattering varying from place to place. The longest wavelengths are constrained by seismic tomography but separating thermal from compositional contributions has major uncertainties. This section reviews the dynamic theory about the expected length scales of heterogeneity.

Mantle heterogeneity is introduced at long wavelengths by subduction (and possibly, primordial layering). If subducted material segregates into a layer, then this acts to increase its wavelength, but if it remains in the convective flow, then it is progressively stretched until it reaches small enough length scales to be homogenized by diffusion. Thus, as discussed by Olson et al. (1984a), the spectrum of mantle heterogeneity undergoing stirring is expected to display three regions with increasing wave numbers (decreasing wavelength): a 'source' region corresponding to the spectral characteristics of the source, a 'cascade' region in which the heterogeneities are being stretched to shorter wavelengths, and a 'dissipation' region, in which chemical diffusion acts to eliminate heterogeneity at the shortest scales. It is useful to consider the spectrum that is obtained in statistical equilibrium, in which heterogeneity is introduced at a constant rate and the total spectrum does not change with time. This depends on the stretching regime in the cascade region: For laminar stirring, the wave number k of an individual heterogeneity increases linearly with time, which leads to a flat ('white') spectrum for all heterogeneities in statistical equilibrium (Olson et al., 1984a). In contrast, for turbulent stirring, the wave vector of an individual heterogeneity increases exponentially with time, that is, the rate of increase is proportional to its present value, which leads to a statistical equilibrium spectrum that scales as $1/k$. This at first seems inconsistent with the $-5/3$ slope obtained by Metcalfe et al. (1995) for a kinematic chaotic stirring of a laboratory fluid, but as this system was not in statistical equilibrium, the slope reflects a transient regime.

Geochemical observations, as discussed in **Section 7.12.1.7**, can be compared to these theoretical slopes. The early analysis of Gurnis (1986b) using $^{87}Sr/^{86}Sr$ data from MORB and OIB appeared to be consistent with a 'white' laminar mixing spectrum, but the more highly constrained spectrum of isotope ratios along the Mid-Atlantic Ridge obtained by Agranier et al. (2005) instead indicates the turbulent, $1/k$ regime, consistent with expectations from mantle mixing theory discussed earlier.

Such theory also successfully predicts the smaller scales of heterogeneity observed in peridotite massifs in Beni Bousera (Kellogg and Turcotte, 1990). They derive that the probability of finding a strip of material larger than size d is given by $1/d(1-F(d))$, which, when evaluated for scales between 1 and 100 cm, gives a slope close to -1, indicating that the $(1/d)$ term is dominating and $(1-F(d))$ does not change very much over this range. Field samples match this slope reasonably well.

In any case, it is clear that because thermal heterogeneity diffuses rapidly, whereas chemical heterogeneity does not, short wavelength mantle heterogeneity is dominantly chemical in origin, while thermal heterogeneity should be important, and perhaps dominant, at long wavelengths, although recent seismological studies argue for a large chemical component (Deschamps et al., 2007; Mosca et al., 2012; Trampert et al., 2004). These expectations are confirmed in the thermochemical convection calculations of Nakagawa and Tackley (2005b), in which the slope of the chemical heterogeneity at long wavelengths also appears to be consistent with a $1/k$ scaling, although they do not study how this scales at short wavelengths.

7.12.3 Mantle Convection with Active (Buoyant) Chemical Heterogeneity

This section considers thermochemical mantle convection in which the composition affects density. Studies have typically considered what happens with a preexisting dense layer or the development of a layer over time by settling of dense material or reactions with the core. Because the chemical diffusivity in the mantle is orders of magnitude lower than thermal diffusivity, it is typically assumed that the chemical field is nondiffusive, although if the field is being treated as a continuous field, then some numerical diffusion is typically introduced in order to stabilize the advection. The Lewis number gives the ratio of chemical to thermal diffusivity:

$$\mathrm{Le} = \frac{\kappa_T}{\kappa_C} \qquad [37]$$

Its value is quoted as $\sim 10^8$ by Hansen and Yuen (1989), but the chemical diffusivity quoted by Kellogg and Turcotte (1986) would lead to $\mathrm{Le} = 10^{13}$.

Attention has generally focused on two cases: compositional layering at the 660 km interface and a relatively thin composition layer above the CMB corresponding to D″.

7.12.3.1 Stability and Dynamics of Chemical Layering in a Convecting Mantle

7.12.3.1.1 The balance between chemical and thermal buoyancy

If a layer of dense material is present, the parameter that most strongly influences its dynamics and evolution is the ratio of chemical density contrast to thermal density contrast, that is, the buoyancy ratio, here written as B:

$$B = \frac{\Delta\rho_C}{\Delta\rho_T} = \frac{\Delta\rho_C}{\rho_0 \alpha \Delta T} \qquad [38]$$

where $\Delta\rho_C$ is the density contrast due to composition, $\Delta\rho_T$ is that due to temperature, ρ_0 is the density, and ΔT is the temperature contrast. If ΔT represents the total thermal contrast available in the system, then it has been found that the layering displays long-term stability with a fairly flat interface if B is greater than some critical value of order 1 and undergoes either large-topography layering or various shorter-lived behaviors that end in mixing and whole-mantle convection, if $B < 1$ (Davaille, 1999a; Le Bars and Davaille, 2004a; Montague and Kellogg, 2000; Montague et al., 1998; Olson and Kincaid, 1991; Richter and Johnson, 1974; Richter and McKenzie, 1981). The behaviors of short-lived modes were systematically mapped out as a function of B, viscosity contrast, and layer

thickness using laboratory studies (Davaille, 1999a,b; Le Bars and Davaille, 2004a,b) supplemented by mathematical analysis (Le Bars and Davaille, 2002) and are generally consistent with earlier laboratory and numerical studies, which are as follows and illustrated in the domain diagram of **Figure 6**:

(i) $B > 1$ stable stratification, both layers convecting with slow entrainment
(ii) $B > 1$ but thin, nonconvecting lower layer
(iii) $1 > B > 0.5$: layered but with large topography
(iv) $0.3 < B < 0.5$, layer thickness between 0.2 and 0.7: doming, the number of pulsations increases rapidly with viscosity contrast
(v) $B < 0.2$: rapid overturn and mixing of layers

If the two layers initially have the same temperature, then additional transient behavior occurs as a temperature difference develops across the layers, thereby decreasing the total (i.e., thermal + chemical) density contrast. In this case, layers that should be unstable according to the previously mentioned criteria may initially be stable, as in the calculations of Samuel and Farnetani (2003). B may decrease with time due to the entrainment of each layer into the other layer, as discussed in the succeeding text. Indeed, in laboratory experiments of Le Bars and Davaille (2004b) (**Figure 2**), the experiment evolved through several different regimes as the density difference between the layers decreased. Another initial condition that causes transient behavior is a linear compositional gradient throughout the domain and constant temperature, as studied by Hansen and Yuen (2000), who found that after some time, the system developed a stable but large-topography layer. After that, the evolution discussed in the remainder of this section applies.

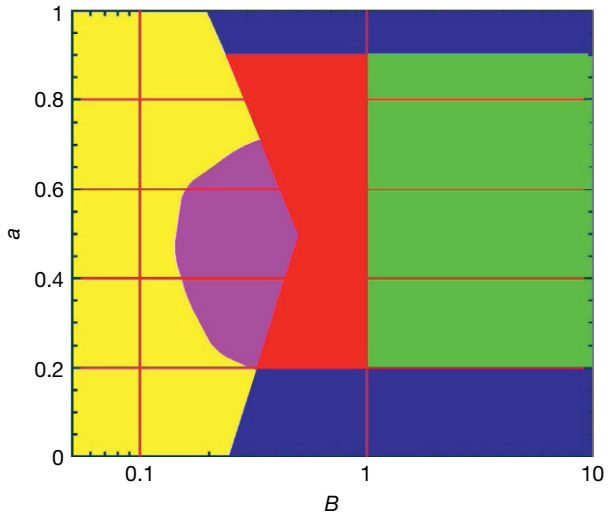


Figure 6 Domain diagram for the behavior of thermochemical convection as a function of buoyancy ratio B and initial depth of dense layer a, as determined by laboratory experiments (Davaille, 1999b; Davaille et al., 2002; Le Bars and Davaille, 2002, 2004a,b). Green, flat layers; red, large-topography layers; pink, transient oscillating domes; blue, thin noninternally convecting layer progressively entrained by thermochemical plumes; yellow, rapid overturn. Adapted from Le Bars M and Davaille A (2004).Whole layer convection in a heterogeneous planetary mantle. *Journal of Geophysical Research* 109: B03403, http://dx.doi.org/03410.01029/02003JB002617.

In the mantle, the relevant physical properties (α, ρ_0) vary with depth, and the previously mentioned criteria are most valid if values of the physical properties close to the chemical boundary area are considered, although some studies have used, for example, reference or 'surface' values and thus require careful interpretation. Scaling to the Earth, the appropriate ΔT is the superadiabatic temperature drop between surface and CMB, which might be 2500 K. For layering at around 660 km depth, $\alpha \sim 2.5 \times 10^{-5}$, so the critical $\Delta\rho_C/\rho_0 \sim 6.25\%$, whereas at the CMB, $\alpha \sim 1.0 \times 10^{-5}$, so $\Delta\rho_C/\rho_0 \sim 2.5\%$. These are rough estimates, as it has been found that other characteristics such as the viscosity contrast between the two layers (a combination of intrinsic viscosity contrast and temperature-dependent contrast), the thickness of the dense layer, and the thermal conductivity of the dense layer also play a role, as will be discussed later.

The detailed dynamics have been investigated variously by laboratory and numerical experiments with some mathematical analyses, with laboratory experiments having the advantage of physical correctness, which is particularly important because entrainment rates are difficult to accurately measure numerically (Tackley and King, 2003; van Keken et al., 1997), whereas numerical experiments have the advantage that the variation of physical properties (viscosity, thermal expansivity, and conductivity) with pressure and temperature can be more realistically ascertained and full information about the various fields is available for analysis. Experiments (both numerical and laboratory) with layering are often initialized in a way that is far from a steady state (e.g., temperature is isothermal instead of higher in the lower layer), which results in a transient behavior. Furthermore, layered experiments are intrinsically transient in that the layering evolves with time, and this may result in the final state retaining a memory of the initial conditions. Further discussion of laboratory experiments can be found in **Chapter 7.03**.

7.12.3.1.2 Initially stable layering: Pattern and dynamics

7.12.3.1.2.1 Flat, thick, internally convecting layers

For thick, convecting layers, layered convection occurs and a TBL builds up over the interface, with heat transport accomplished by thermal diffusion, substantially reducing the global heat flow and affecting the Earth's thermal evolution (Montague and Kellogg, 2000). For layers of approximately equal thickness and viscosity, as in the laboratory and numerical experiments of Richter and McKenzie (1981), the temperature difference between the layers is $\Delta T/2$ (Richter and McKenzie, 1981), but this varies as the viscosity contrast and thicknesses are changed. The pattern of convection in the upper and lower layers is coupled. When the layers are of similar viscosity, mechanical coupling dominates such that upwellings in one layer match with downwellings in the other layer (Cserepes and Rabinowicz, 1985; Richter and McKenzie, 1981). When there is a large viscosity contrast exceeding a factor of ~100 (Cserepes and Rabinowicz, 1985), however, thermal coupling dominates, in which upwellings in the lower layer underlie upwellings in the upper layer. Intermediate viscosity contrasts, such as 25–30, result in a mixture of thermal and mechanical (viscous) coupling, as show by 3-D calculations (Cserepes et al., 1988; Glatzmaier and Schubert, 1993). In this mixed regime, the flow direction of rolls in the upper layer was observed to become perpendicular to those in

the lower layer by Cserepes et al. (1988). Entrainment occurs at cusps in the interface, as will be discussed later.

If one layer is substantially more viscous, the form of upwellings or downwellings forming at the interface is also different (Davaille, 1999b), with 2-D sheets forming in the more viscous layer but 3-D plumes forming in the less viscous layer, which have different entrainment rates as will be discussed later.

7.12.3.1.2.2 Thin, nonconvecting lower layer

In this case, the lower layer is embedded within the lower TBL of the main convection, and the dense material gets swept by the convection, forming a spoke pattern of ridges on top, with plumes at the junctions of the ridges, and entrainment takes place there, as shown by 3-D numerical experiments (Tackley, 1998) and laboratory experiments (Davaille et al., 2002; Olson and Kincaid, 1991). The numerical experiments show that this dense layer acts like a rigid boundary, reducing the horizontal wavelength of convective planform relative to a free-slip lower boundary (Tackley, 1998), and that if the layer internally convects (Montague and Kellogg, 2000; Montague et al., 1998), as would be the case if its viscosity is low enough (Schott and Yuen, 2004; Solomatov and Moresi, 2002), small-wavelength thermal structure is generated within the layer. These numerical experiments also show that even in 100% internally heated cases, entrainment exists at cusps anticorrelated with where the downwellings are, and larger amounts of internal heating reduce the buoyancy ratio needed for stability. Temperature-dependent viscosity decreases the density contrast needed for stable layering (Schott et al., 2002; Tackley, 1998) and, when the layer is marginally stable, can lead to complicated schlieren structures of entrained material and complicated layer topography (Schott et al., 2002). For reasonable Earth parameters, Tackley (1998) found for an internally heated case that ~1% density contrast was sufficient to maintain a layer over reasonable timescales.

7.12.3.1.2.3 Thick, large-topography lower layer or isolated piles

With a fairly low buoyancy ratio like 0.6 (Montague et al., 1998), the topography on the layer interface is large (Davaille, 1999a; Kellogg et al., 1999), which may result in discontinuous 'piles' or 'hills' if the layer thickness is less than the topography. In 2-D calculations at low Rayleigh number, this results in wide stable 'hills' below the upwelling, which cause strong (approximately several km) negative topography on the CMB (Davies and Gurnis, 1986; Hansen and Yuen, 1989), with temperature-dependent viscosity decreasing the magnitude of this negative topography (Hansen and Yuen, 1989). The topography on a layer interface decreases with increasing Rayleigh number, approximately as $Ra^{-1/3}$ (Davies and Gurnis, 1986; Tackley, 1998), which predicts relatively little topography at the Earth's convective vigor (Montague and Kellogg, 2000). If, however, the increase of viscosity and thermal conductivity and the decrease of thermal expansivity with depth are taken into account, they lead to a low effective *Ra* for the deep mantle, so large interface fluctuations can still occur, as shown in 3-D (Tackley, 1998, 2002). It has been argued that such structures may account for the superplumes observed in seismic tomography, discussed in **Section 7.12.1.8**. An important point shown by van Thienen et al. (2005a) is that any layering around 660 km induced by the postspinel phase transition makes the layer more stable over long time periods, presumably because it reduces the temperature drop across the layer and the convective velocities.

Regarding the three-dimensional structure of a strongly undulating layer, Cartesian calculations (Tackley, 1998, 2002) that were either Boussinesq or compressible with constant, depth-dependent, or depth- and temperature-dependent viscosity, found that the long-term structure of dense piles tends to be ridges/spokes. In spherical geometry Boussinesq calculations (Oldham and Davies, 2004), there is a greater tendency for the piles to be isolated. A viscosity contrast between the layers makes a major difference to the planform: McNamara and Zhong (2004b) showed that temperature-dependent viscosity leads to linear piles that are swept around and spread through the entire lower mantle, whereas composition-dependent viscosity in which the dense material is more viscous leads to isolated round 'superplumes,' particularly with contrasts as large as factor 500. This seems to be consistent with the 'domes' in the laboratory results of Davaille (1999a), except that they also observe domes when the dense layer is less viscous. The reconciliation of these is that such domes are transient features – for long-lived features, domes are only observed when the lower layer is more viscous (McNamara and Zhong, 2004b; see also **Chapter 7.11**).

Comparison of structures obtained from convection models with structures imaged by seismic tomography provides important constraints. Deschamps and Tackley (2008, 2009) performed a systematic suite of thermochemical convection models running for billions of years and compared the results to probabilistic seismic tomography of Trampert et al. (2004). Convection models that matched seismic constraints had a moderate compositional density difference of 90–150 kg m^{-3} (leading to large-topography piles), a large thermal viscosity contrast, and at 660 km depth a viscosity jump by factor 30 and an endothermic phase transition with a Clapeyron slope of -1.5 to -3 MPa K^{-1}.

7.12.3.1.2.4 Long-term fixity of piles?

Based on the observation that present-day hot spots and the reconstructed locations of large igneous provinces (LIPs), both assumed to be caused by plumes from the CMB, often coincide with the edges of the large seismically slow regions above the CMB, particularly under Africa (Burke and Torsvik, 2004; Thorne et al., 2004; Torsvik et al., 2006), it has been proposed that plumes are preferentially generated edges of dense piles that have remained in a similar position for the last ~300 Ma (Burke et al., 2008; Torsvik et al., 2008, 2010). There has even been a proposal that the large-scale structure of the deep mantle has been stable for billions of years (Dziewonski et al., 2010). Important questions are thus (i) whether such long-term stability is dynamically feasible and (ii) why do plumes come from the edges, rather than centers, of the piles? Here, (i) is considered; plumes are discussed in **Section 7.12.3.1.3**. In all laboratory or numerical experiments of thermochemical convection in which the flow is time-dependent, piles of dense material are swept around by the large-scale flow, sometimes being split by new downwelling slabs or merging (e.g., Davaille et al., 2002; Hansen and Yuen, 1989; McNamara et al., 2010; Tan et al., 2011). Thus, stability for billions of years

seems unlikely. However, even in time-dependent, chaotic convection, there are often features that remain in a similar position for some time, and the proposed 300 Ma stability is less than one 'overturn time' of mantle convection. Indeed, in many published numerical or laboratory studies, it is possible to find thermochemical piles that are stable for 100s Ma (e.g., Figure 6 in Tan et al., 2011). Thus, a special explanation for a few 100s Ma fixity of dense piles is probably not required (see also **Chapter 7.11**).

7.12.3.1.2.5 Models driven by plate reconstructions

Several studies have focused on modeling the last ~200 million years of the Earth's evolution by applying plate reconstructions as an upper velocity boundary condition on 3-D spherical thermochemical convection models, following earlier studies for purely thermal convection (e.g., Bunge et al., 1998). McNamara and Zhong (2005) found that thermochemical structures similar to those seismologically observed under Africa and the Pacific were generated. Specifically, the Earth's subduction history tends to focus dense material into a ridge-like pile beneath Africa and a relatively more rounded pile under the Pacific ocean, which they argued is consistent with seismic observations, a point that was further reinforced by performing synthetic seismic tomography of such models (Bull et al., 2009). Bower et al. (2013) performed similar modeling but with the addition of progressive assimilation to guide the thermal structure of the lithosphere and steer the thermal evolution of slabs in the upper mantle and also a different (higher) bulk modulus for the dense material. They obtained deep mantle thermochemical structures that are broadly similar to those of seismic tomography and of McNamara and Zhong (2005), found that spreading slabs can cause steep sides and plumes at the edges of dense piles, and found that having a higher bulk modulus for the dense material was not necessary. While thermochemical structures are typically invoked, there is an active debate as to whether purely thermal convection can generate structures resembling seismic tomography (Davies et al., 2012).

7.12.3.1.2.6 Steep sides

In some places on Earth, sharp near-vertical sides to the large low-velocity provinces are observed seismically (e.g., He and Wen, 2009; Ni et al., 2002, 2005; Wang and Wen, 2007). Two possible causes of these are (i) slabs spreading above the CMB 'bulldozing' the dense material, which is a transient effect, and (ii) the dense material having a higher bulk modulus than the surrounding mantle. Transient steep sides (timescale ~100s Ma) caused by 'bulldozing' were observed in laboratory experiments (Davaille, 1999a) and numerical experiments (Deschamps and Tackley, 2008, 2009) of free convection and numerical experiments with imposed plate motions (Bower et al., 2013; Bull et al., 2009; Lassak et al., 2010; Steinberger and Torsvik, 2012). Long-term stable steep-sided thermochemical piles were investigated by Tan and Gurnis (2005) and Tan and Gurnis (2007), who found that an increase in chemical density contrast with height above the CMB is necessary. This is because the thermal expansivity increases with height above the CMB, making the top of the piles less stable unless the chemical buoyancy contrast also increases. Whether MORB has this property is uncertain, as discussed in **Section 7.12.3.3.3**. Such decreasing MORB density contrast with depth has also been included in some numerical simulations in which a layer accumulates over time by segregation of MORB rather than inserted a priori (Nakagawa and Tackley, 2005a, 2010, 2011, 2013; Xie and Tackley, 2004a,b), and steep-sided thermochemical structures are observed in some of those calculations. The rate at which thermal expansivity changes with depth is very small in the deep mantle (and is exaggerated in by Tan and Gurnis (2005) and Tan and Gurnis (2007) by assuming a linear profile), so the needed difference in compressibility is small.

7.12.3.1.2.7 CMB topography

If piles have a low viscosity, then they do not depress the CMB as earlier models indicated (Hansen and Yuen, 1989), but instead cause relatively flat CMB topography under themselves, with lower CMB topography occurring under slab graveyards (Lassak et al., 2007, 2010).

7.12.3.1.3 Influence of a dense layer on plume dynamics

7.12.3.1.3.1 Temperature

The presence of a dense layer reduces the temperature anomaly of plumes (Farnetani, 1997; Zhong, 2006), which may help to explain why the inferred temperature anomaly of plumes close to the lithosphere is much lower than the estimated temperature drop across the lower TBL for whole-mantle convection.

7.12.3.1.3.2 Stability

Laboratory experiments find that plumes rising from cusps of a dense layer are stable over longer time periods than plumes rising from the lower TBL of whole-mantle convection. In particular, Davaille et al. (2002) found that the dense layer acts as a 'floating anchor,' reducing the lateral migration of plumes, which then have a scaled drift velocity of 1–2 mm year^{-1} – larger than simple scalings predict for thermal plumes. In addition, the dense layer enables them to survive for 100s My. The plume spacing is proportional to the stratified boundary layer thickness. In the laboratory experiments of Jellinek and Manga (2002), the dense, low-viscosity material forms a network of ridges (embayment and divides), and they delineate the trade-off between viscosity contrast and layer topography in the stabilization of plumes. They found that if the dense material has a similar viscosity, the topography required to stabilize plumes is greater than two to three times the TBL thickness (to overcome viscous stresses), but this required topography decreases as the layer viscosity decreases, because the lower layer then acts as a free-slip boundary. Indeed, with >2.5 orders of magnitude viscosity contrast, only minimal topography is needed. With no dense layer present, upwellings are in the form of intermittent thermals.

This laboratory finding that a dense layer stabilizes plumes is seemingly contradicted by numerical experiments (McNamara and Zhong, 2004a). In these numerical experiments, plume stability was quantified as the ratio of horizontal plume velocity to average surface velocity. McNamara and Zhong (2004a) found this to be about the same in layered calculations as it is in isochemical calculations with a rigid lower boundary, regardless of the viscosity contrast of the dense material. They also found that layer topography does not influence fixity, but can lead to longer-lived plume

conduits that follow the topographic peaks around. The reconciliation of these apparently contradictory findings is not clear. One possibility is that the dense layer reduces the convective vigor and that the laboratory experiments did not take this into account. Another is that there is a difference in the dynamics of downwellings (e.g., they might be stronger in the numerical experiments), which push the lower layer and associated hot plumes around. More study is needed to reconcile laboratory and numerical results on this issue.

7.12.3.1.3.3 Plumes from the edges of piles?

Observations suggest that in the Earth, present-day plumes (Thorne et al., 2004) and historical LIPs inferred to be caused by plumes (Burke et al., 2008; Burke and Torsvik, 2004; Torsvik et al., 2006, 2008, 2010) tend to come from the edges of inferred chemically dense piles (i.e., low seismic-velocity provinces), rather than the tops of them. In detail, this appears to be true under Africa, but under the Pacific, plumes also emerge from the tops of the inferred dense pile. In contrast, in most laboratory and numerical experiments of thermochemical convection, plumes typically form from the tops of dense piles or ridges, rather than their edges. Recent numerical experiments indicate that key mechanisms for generating plumes from the edges of piles are slabs and steep sides. Slabs reaching the CMB tend to induce plumes at their edges, even in the absence of chemical layers (e.g., Tan et al., 2002). When chemical piles are present, slabs can 'bulldoze' the dense material to create ridges at their edges, which promotes plume formation. The latter can be seen in thermochemical convection calculations with imposed historical plate motions (Bower et al., 2013; Bull et al., 2009; Lassak et al., 2010; Steinberger and Torsvik, 2012), in which cusps from which plumes can form appear at the edges of the Pacific and African low-velocity provinces. In regional calculations of compositionally stratified slabs reaching the CMB (Tackley, 2011), plumes also tend to form at the edges of regions of dense material, which are also the edges of the slabs. Steep edges can be generated by the 'bulldozer' effect of slabs, potentially helped by the dense material having a higher bulk modulus (Tan and Gurnis, 2005, 2007), although the latter is not sufficient to guarantee edge plumes as the plumes in Tan and Gurnis (2005, 2007), in which there are no cold downwellings, form at the tops of the piles. The most successful models to obtain plumes from the edges of piles without any imposed motions are those of Tan et al. (2011), who performed free convection models of thermochemical convection in 3-D spherical geometry with a moderate viscosity contrast and obtained two to three steep-sided piles with plumes arising mostly from their edges. Another effect that can suppress plumes from the tops of piles is for them to have a compositional gradient rather than a sharp compositional boundary (Tackley, 2011). Layers that build up by segregation of subducted MORB (e.g., Brandenburg et al., 2008; Nakagawa et al., 2010) tend to be topped by such compositional gradients rather than a sharp boundary.

7.12.3.1.3.4 Starting plumes

Despite the effect that a dense layer can have on increasing plume stability as discussed earlier, some studies have found that for starting plumes, the presence of a dense layer can add complexity and modes of time dependence that are not present with purely thermal plumes. Lin and Van Keken (2006a,b) numerically investigated axisymmetric models set up to produce a plume at the axis, various temperature dependences of viscosity, and a dense layer with various thicknesses, buoyancy ratios, and viscosity contrasts. They found that entrainment commonly becomes a transient process for models with variable viscosity. In models where a large fraction of the dense layer was entrained, variable viscosity could lead to secondary instabilities developing after the initial plume head, whereas in models with only a thin filament entrained, variable viscosity could lead to an initial pulse of entrainment. In general, the dense material led to a wider range of plume morphologies than observed with purely thermal plumes. As with results discussed elsewhere, they found that decreasing the viscosity of the layer made it more stable.

Even greater morphological variability and complexity were found in the 3-D models of Farnetani and Samuel (2005), in which thermochemical plume formation was modeled in a compressible mantle with no imposed symmetry, an initially 200 km thick layer with buoyancy ratio $B=0.1$, and T- and depth-dependent viscosity. Plume morphologies were much less regular than 'classical' plumes and included plumes with blob-like or irregular heads, tilted plumes, ~1000 km wide 'megaplumes,' broad plumes without heads, and 'classical' head plus narrow tail plumes, often with several types coexisting. The buoyancy flux of conduits varied widely, and some structures ponded under the 660 km discontinuity. Plume conduits were found to be laterally heterogeneous (i.e., without concentric zonation), but as with previous results (Farnetani et al., 2002), the only material hot enough to melt underneath the lithosphere is from the source region. If the dense material has a different compressibility, this can lead to additional complexity such as plume heads ponding at a certain depth (Samuel and Bercovici, 2006).

Laboratory experiments of thermochemical starting plumes (Kumagai et al., 2008) confirmed the conclusions from the previously mentioned numerical experiments: thermochemical plumes may not have the classical plume structure but can also be fat and patchy, and they may also exhibit more complexity and time dependence than purely thermal plumes.

7.12.3.1.4 Entrainment of surrounding material by mantle plumes

As discussed earlier, plumes clearly entrain material from a dense layer at their base. A long-standing question is how much this composition is diluted by material entrained from the upper layer as they rise. The initial plume head and subsequent conduit have different structure and so are expected to be different in their entrainment characteristics.

Depending on the viscosity contrast, rising hot diapirs, analogous to plume heads, thermally entrain surrounding material as they rise, forming a folded structure (Griffiths, 1986). It was demonstrated that this structure also occurs in plume conduits if they are sheared by the surrounding material (Richards and Griffiths, 1989), to such an extent that "small plumes or narrow plume conduits may consist of only a small fraction of original source material" (Richards and Griffiths, 1989). Hauri et al. (1994) presented analytic solutions for plume conduits, finding that they entrain significant amounts

of ambient mantle due to lateral thermal conduction from the plume. They found a huge range of possible entrainment, depending on the buoyancy flux, with most of the entrained material coming from the lower half of the layer they are passing through.

Although plumes may entrain a significant amount of surrounding material, Farnetani and Richards (1995) found that the entrained material does not contribute significantly to the melt produced by the plume, with >90% of the primary plume magmas being composed of primary plume material. The first-order reason for this is that the entrained material has a much lower temperature than the core of the original plume, and so is much less likely to melt underneath the lithosphere. Reinforcing and extending this finding, Farnetani et al. (2002) analyzed entrainment by a thermal plume with temperature-dependent viscosity contrast 100 and found that the material that melts is from the TBL at the base of the mantle. The material in the plume head was found to be much more extensively stretched and stirred than that in the plume conduit, implying greater homogenization of geochemical anomalies. At that viscosity contrast, the plume head did not entrain surrounding material, although a substantial amount of lower mantle material was advected upward with the plume into the upper mantle. Subsequent models (Farnetani and Hofmann, 2009, 2010) reinforced this conclusion.

The influence of plume entrainment on trace-element signatures is discussed in **Section 7.12.4.4**.

7.12.3.1.5 Ultra-low-velocity zones

A number of geographically localized and vertically thin (~10 s km) zones of very low seismic velocity (5–10% lower than average) directly above the CMB have been detected since the mid-1990s (Garnero and Helmberger, 1995, 1996). Their locations are statistically correlated with hot-spot locations (Williams et al., 1998) and with the edges of large, seismically slow regions (McNamara et al., 2010) that are commonly inferred to be compositionally distinct, as discussed in **Section 7.12.1.5.4**. Two classes of explanation have been proposed: (i) 10–30% partial melt and (ii) some type of iron-enriched solid. Dynamic aspects of these possibilities are reviewed in the succeeding text (see also **Chapter 7.11**).

If ULVZs are caused by 10–30% partial melt (Garnero, 2004; Lay et al., 2004; Revenaugh and Meyer, 1997; Vidale and Hedlin, 1998; Wen and Helmberger, 1998; Williams and Garnero, 1996), a key question is why the melt stays in the solid matrix rather than draining into a pure melt layer above the CMB (if it is denser than the solid) or rising into the colder mantle above and solidifying (if it is buoyant). Dynamic calculations of the deepest mantle including two-phase flow indicated a very narrow parameter range in which melt would remain in the solid to form ULVZ-like features (Hernlund and Tackley, 2007): the melt must be slightly denser than the solid and the permeability curiously very low. These calculations, however, used a simplified treatment of melt migration in which the driving force for melt percolation was purely gravity acting on the density difference between solid and liquid. Hernlund and Jellinek (2010) showed that when dynamic pressure gradients associated with deformation of the solid matrix by surrounding flow are also included, dynamic pressure gradients can dominate gravitational settling and allow the melt to stay in suspension in the solid.

Various solid compositions are also plausible candidates for explaining ULVZ seismic properties and may match better than partial melt. These include iron-rich postperovskite (Caracas and Cohen, 2005; Mao et al., 2006; Stackhouse et al., 2006; Stackhouse and Brodholt, 2008), a mixture of iron-enriched (Mg,Fe)O magnesiowüstite and ambient mantle (Bower et al., 2011; Wicks et al., 2010), and subducted banded iron formations (Dobson and Brodholt, 2005). The origin of iron-enriched material might be the final product of crystallization of a basal magma ocean (BMO) (Labrosse et al., 2007; Nomura et al., 2011; **Section 7.12.5.2.5**) or chemical reactions with the core (**Section 7.12.1.5.4**). It has also been suggested that ULVZs could be caused by accumulated silicate sediments from the core (Buffett et al., 2000).

The dynamic feasibility of the 'dense solid' explanation was demonstrated by McNamara et al. (2010) using calculations of thermochemical convection with two types of dense material, one representing large-scale piles and the other small-scale features with the density seismically inferred for ULVZs. They showed that small volumes of solid ULVZ material can be locally elevated despite its high density and are swept around by the convection to the boundaries of large compositional piles, which periodically break apart and merge together in response to changes in downwelling patterns. When the large piles merge, some ULVZ material is initially located inside the merged pile, in accordance with recent seismological observations (Thorne et al., 2013). The detailed morphology and seismic signature of isolated solid ULVZs was investigated by Bower et al. (2011), who identified a parameter range in which they resemble seismically observed structures (see also **Chapter 7.11**).

7.12.3.2 Long-Term Entrainment of Stable Layers

In a system with stable compositional layering, the system nevertheless evolves with time because the thermal convection in each layer causes entrainment of material from the other layer, in the form of thin schlieren, thereby diminishing the density contrast (Olson, 1984). There are two possible outcomes to this evolution: (i) If the entrainment is asymmetrical (typically due to the lower layer being thinner than the other, e.g., Samuel and Farnetani, 2003), then one layer may eventually disappear due to complete entrainment into the other layer. Indeed, it is likely that in the long term, there is substantial entrainment of any dense material residing above the CMB, which may be balanced by the addition of material by segregation or core reactions, as will be discussed later. (ii) The density contrast decreases to the point where the layering is no longer stable and large-scale overturn and stirring occur. Numerical calculations and laboratory experiments have been useful in elucidating the details of these processes, while mathematical analysis has been useful in developing scaling laws for entrainment.

7.12.3.2.1 Comparison of entrainment laws and estimates

Several authors have obtained expressions for the entrainment rate as a function of parameters such as density contrast, viscosity contrast, and so forth. Studies usually focus on one of

two situations: a thin layer at the base of the mantle or two layers of substantial thickness. While the same laws might apply, the former case obviously has much greater asymmetry and the lower layer might not be convecting.

7.12.3.2.1.1 Laboratory experiments

From studying the entrainment in laboratory experiments with two layers of equal depths, Olson (1984) found that the mixing rate of the two layers, defined as the rate of change of their density divided by the density difference (i.e., the fractional volume flux), scales as

$$\dot{M}/\dot{\varepsilon} \approx 0.046 Ri^{-1} \quad [39]$$

where the Richardson $Ri = \Delta\rho g D/\tau$ with τ being the average viscous stress within each layer. Eliminating Ri leads to

$$\dot{M}/\dot{\varepsilon} \approx 0.046 \frac{\tau}{\Delta\rho g D} \quad [40]$$

For two thick stable layers where the lower layer is more viscous by a factor of $1.0–6.4 \times 10^4$, Davaille (1999b) found that the scaling for entrainment by linear sheets (typically found in the more viscous layer) is different from that for axisymmetric plumes (found in the less viscous layer). For entrainment by linear sheets,

$$Q_{21} = C_1 \frac{\kappa h_0}{B} Ra^{1/5} \quad [41]$$

whereas for columnar plumes,

$$Q_{12} = C_2 \frac{\kappa h_0}{B^2} Ra^{1/3} \quad [42]$$

where Q is flux of material, h_0 is initial thickness of lower layer, κ is thermal diffusivity, and C_1 and C_2 are two experimentally determined constants. The 'plume' law has the same $1/B^2$ form as the one proposed by Sleep (1988) (discussed in the succeeding text). When the lower layer is substantially more viscous than the upper layer, it is difficult for the upper layer to entrain it, requiring an additional factor (Davaille, 1999a):

$$Q = C_1 \frac{\kappa H}{B^2} Ra^{1/3} \frac{1}{1 + \Delta\eta/B} \quad [43]$$

where $\Delta\eta$ is the viscosity contrast. This additional factor applies for $\Delta\eta > 1$, that is, for entrainment of the more viscous layer into the less viscous layer. If the layer being entrained is substantially less viscous than the entraining layer, then Q could also be reduced due to draining of the less viscous material back into the layer it came from, but this was not observed by Davaille (1999a) and Davaille et al. (2002) at $\Delta\eta$ as low as 0.01. This entrainment law was subsequently verified for the case of a thin layer with plume entrainment (Davaille et al., 2002). Depending on parameters, a mantle plume would entrain chemically dense material at a rate of 0.1–15% of its total volume flux. Davaille et al. (2002) presented graphs of layer survival time versus thickness and density contrast, based on thermal expansivity $10^{-5}\,K^{-1}$ and 'temperature heterogeneity' of 500 K. As an example, entrainment of 100 km over 4.5 Gy occurs with a 3.5% density contrast if the layers are of the same viscosity or ~2% density contrast if the dense material is ten times more viscous.

Several other authors have conducted laboratory experiments to study entrainment and long-term evolution. In experiments with an initial layer thickness of 0.2 (600 km) and similar viscosity materials, Gonnermann et al. (2002) found vigorous convection in both layers and entrainment consistent with the scaling laws of Davaille (1999a). They estimated that an initially 300 km thick layer can survive over the Earth's history if it has a 2% density contrast, but if initially thicker, then it can survive if it has a <2% initial density contrast. In some cases, the dynamics changed from stable layering to doming.

The strong influence of viscosity contrast on entrainment of a thin layer was demonstrated by Namiki (2003) using viscosity ratios from 0.1 to 400. It was found that when the lower layer is more viscous, it grows in size due to asymmetrical entrainment of the upper layer. When the lower layer is less viscous, an interfacial layer developed and the volumes of the layers did not change because plumes entrained the interfacial layer, not the lower layer. This interfacial layer is a novel finding not observed in any other mantle-related studies, so it is worth considering what causes it. The layer might develop because of incomplete entrainment of the lower, less viscous layer by the upper layer due to reduced viscous coupling. It might form as a result of double-diffusive convection, which helps to transport material across the boundary (M. Manga, personal communication). If the latter explanation is correct, it would not form in Earth because the chemical diffusion rate is very low. In any case, by considering stresses (as in Olson, 1984), Namiki (2003) derived an entrainment law that is consistent with those discussed earlier:

$$\dot{E}_i \sim \frac{\kappa_i}{B_i^2 \delta_{th_i}} \quad [44]$$

where $\dot{E}_i$ is the entrainment rate of layer i and δ is the TBL thickness, related to the Rayleigh number as $\delta_{th} \sim 5 L_i Ra^{-1/3}$ leading to an overall scaling of

$$\dot{E}_i \sim 0.2 \frac{\kappa_i}{B_i^2 L_i} Ra^{1/3} \quad [45]$$

In this parameterization, the effect of a viscosity difference is to alter the ΔT for each layer, hence B and Ra. The survival time of a 200 km thick layer was calculated as a function of its density difference, assuming the total $\Delta T = 4000$ K, the viscosity of D″ is 10^{19} Pa s, the viscosity of the mantle is 10^{21} or 10^{22} Pa s, and thermal expansivity is $2 \times 10^{-5}\,K^{-1}$. It was estimated that a density difference of at least 300 kg m^{-2} (5.5%) is needed for the layer to survive over geologic time (ignoring the possible presence of an interfacial layer). With a more realistic thermal expansivity of $1 \times 10^{-5}\,K^{-1}$, this would be reduced to 2.75% contrast, comparable with other estimates.

A scaling law that explicitly includes viscosity contrast was proposed by Jellinek and Manga (2002) on the basis of laboratory experiments:

$$q = 0.49 \frac{\kappa}{H} \frac{Ra^{1/3}}{\Delta\eta B} \quad [46]$$

where H is the depth of tank and $\Delta\eta > 1$ is the ratio of upper to lower layer viscosity. This is similar to the expression of Davaille (1999a) given earlier, in the limit of $\Delta\eta \gg B$. It is interesting that in this limit, the dependence on B is $1/B$ rather than $1/B^2$.

7.12.3.2.1.2 Mathematical or numerical approaches

Gradual entrainment of thin chemical layer at the base of the mantle by either 2-D upwelling sheets or axisymmetric plumes was analyzed by Sleep (1988) to determine the fraction of entrained material in the upwelling as a function of chemical density contrast. The analysis did not consider the details of the entrainment process, but rather the carrying capacity of the plume conduit. Thus, a 1-D section across the upwelling was considered, with the entrained material in the form of a thin filament in the center of the upwelling. To determine the thickness of this filament, it was assumed that the thickness adjusts so as to maximize entrainment rate (e.g., if it is too thick, then the upwelling velocity decreases). The models included temperature-dependent viscosity. While a general scaling law for all parameter combinations was not explicitly stated, the general form was found to be

$$F = C/B^2 \qquad [47]$$

where C is a constant and B is the chemical buoyancy ratio. Based on these models, Sleep (1988) estimated that a chemical density contrast of 6% is necessary for a 100 km thick, non-convecting layer to survive over the age of the Earth if the thermal expansivity is $2 \times 10^{-5}\,\mathrm{K}^{-1}$. Subsequent mineral physics data (e.g., Anderson et al., 1992; Chopelas, 1996) indicate a thermal expansivity closer to $1 \times 10^{-5}\,\mathrm{K}^{-1}$ at the CMB, so this estimate of density contrast should be adjusted to 3%. Additionally, if the initial layer thickness were larger, then a higher entrainment rate could be tolerated. In any case, this density contrast corresponds to $B \sim 3$ based on the temperature drop over the lower boundary layer or $B \sim 1.5$ using the temperature drop over the entire system – both substantially larger than that required to stabilize the layer against rapid overturn. Sleep (1988) also considered the case of an internally convecting layer and concluded that the earlier mentioned estimate is still reasonable.

The entrainment process at the base of the plume (not considered by Sleep, 1988) was included in the high-resolution 2-D numerical models of Zhong and Hager (2003). They found that treating the lateral flow at the bottom of the plume makes some significant differences compared to the 1-D analysis of Sleep (1988), with the 1-D model overestimating entrainment by up to more than a factor of 4 compared to their 2-D calculations. They found that the entrainment rate Q is mainly controlled by the buoyancy ratio B and the radius of the thermal plume r_T but insensitive to the thickness of the dense layer, with a relationship

$$Q \sim B^{-2.48} r_T^{3.80} \qquad [48]$$

Because their model was set up to produce a plume with prescribed parameters, they did not obtain a scaling law for a full convective system. They did, however, apply the scaling to estimate the survival time of a 1000 km thick layer and estimated that more than 90% of a 1000 km thick dense layer can survive for 4.5 Gy if the dense material has a net negative buoyancy of ~1%. It is important to note that this is the 'net' buoyancy (thermal + chemical) and not the chemical buoyancy! For a thermal expansivity of $2.5 \times 10^{-5}\,\mathrm{K}^{-1}$ and a dense material 800 K hotter, the chemical density contrast would have to be 3%. This could be reduced by a factor of ~2 using a more realistic deep mantle thermal expansivity.

Recently, Lin and Van Keken (2006a,b) revisited the transient dynamics of a plume forming from a boundary layer with an included dense chemical layer, first studied by Christensen (1984). The dynamics are discussed in **Section 7.12.3.1.3**. Although the short-lived (250 Ma) nature of their simulations was not intended to study long-term entrainment, they did compare short-term entrainment with existing laws and also found some interesting behavior. In particular, they found two effects that are not accounted for in standard entrainment theory: (i) A very thin layer, that is, much thinner than the TBL, could be entrained very rapidly despite having $B = 2$ or 3. (ii) For $B \sim 1$, a new starting plume head caused a 'pulse' in entrainment, bringing into question the applicability of steady-state entrainment laws for a dynamic, time-dependent mantle. As expected, temperature-dependent viscosity resulted in much lower entrainment, for example, with a factor of 1000, even a layer with $B < 0.5$ was stable. Regarding entrainment rate, their results were consistent with $F = C/B^2$ found by Sleep (1988) with $C = 0.07$ and the scaling law of Davaille (1999b) with $C = 1.1$, although there was some deviation from the latter when $B > 1$, which they suggested is due to the effective B and Ra adapting with time in the laboratory experiments. They estimated that a layer with temperature-dependent viscosity can survive if it is thicker than 75 km and its compositional density contrast is larger than 1.5% assuming thermal expansivity is 3×10^{-5}, which would be reduced to ~0.5% with an Earth-like thermal expansivity.

7.12.3.2.1.3 Summary of layer survival estimates

As several studies discussed earlier have estimated the density contrast required for a dense layer to survive over geologic time in the face of continuous entrainment, it is useful to consider whether these estimates are consistent with each other. In general, the published estimates assume a deep mantle thermal expansivity at least a factor of 2 higher than realistic, so in **Table 2**, they have been rescaled if necessary to a thermal expansivity of $1 \times 10^{-5}\,\mathrm{K}^{-1}$. In general, the rescaled estimates are consistent, with thinner layers increasing the required density contrast and a viscosity contrast (either higher or lower) reducing the required density contrast. Perhaps the exception is Lin and Van Keken (2006a), whose density estimate is very low. This could be because they have a larger viscosity contrast than any other study, and this is certainly worth exploring further. Another factor might be that they consider a single plume with $\Delta T = 750$ K rather than a full convective system as in the laboratory studies, and convective downwellings could help to destabilize the layer. On the other hand, Sleep (1988)

Table 2 Scaled estimates of density contrast required for long-term survival

Study	*Thickness (km)*	$\Delta\rho$	η
Sleep (1988)	100	3.0%	1
Davaille et al. (2002)	100	3.5%	1
Davaille et al. (2002)	100	2.0%	10
Gonnermann et al. (2002)	300	2.0%	1
Namiki (2003)	200	2.75%	0.01
Zhong and Hager (2003)	1000	1.5%	1
Lin and Van Keken (2006a)	>75	>0.5%	0.001

and Zhong and Hager (2003) also considered isolated plumes. In summary, this issue requires investigation in a fully convecting system with large viscosity contrast.

7.12.3.3 Evolution of a Mantle with Ongoing Differentiation

7.12.3.3.1 General evolution

So far, this review has focused on the evolution of a system with preexisting layering. However, while the possibility of primordial layering exists, most of the geochemically observed heterogeneity appears to be related to recycled material, so several models have studied the evolution of a mantle in which heterogeneity is continuously introduced over time, usually to simulate the differentiated components in subducted slabs (Christensen, 1989b; Christensen and Hofmann, 1994; Davies, 2002, 2007; Gurnis, 1986a; Huang and Davies, 2007a,b,c; Kameyama et al., 1996; Nakagawa and Tackley, 2005a,b, 2006, 2010, 2011, 2012, 2013, 2014; Nakagawa et al., 2009, 2010, 2012; Ogawa, 1988, 1993, 1997, 2000a, 2003, 2010, 2013; Tackley et al., 2007; van Thienen et al., 2004; Xie and Tackley, 2004a,b). Some of these have already been discussed to a certain extent, for example, in the context of the effect on residence time.

In this section, models in which active heterogeneity is introduced are discussed. Some of the models track trace elements as well as crust residue differentiation; the trace-element aspects of them are discussed in **Section 7.12.4**. One of the main differences between studies is how they treat the secular evolution of the planet, in which the mantle and core cool and heat-producing elements produce less heat. While some studies attempt to directly include this (e.g., Nakagawa and Tackley, 2005a; Ogawa, 2003; Xie and Tackley, 2004b), others have parameterized the faster earlier evolution as a longer run time at a constant convective vigor (Davies, 2002). Other differences are how crustal production is treated (self-consistent melting criterion or assumed thickness crust) and how plate tectonics is treated (kinematic plates or using a yield-stress approach).

While there are substantial variations in the assumptions and results among different calculations, it is possible to generalize the typical evolution. Here and in subsequent discussions of the deep mantle, 'MORB' refers to subducted MORB that actually has a different mineralogy than basalt due to solid–solid phase transitions; for example, it is eclogite at upper mantle pressures:

1. Initial condition: chemically homogeneous.
2. Vigorous early melting phase, characterized by (i) high magmatic heat transport (the so-called 'magmatic heat pipe' mode discussed in **Section 7.12.3.3.2**), (ii) rapid production of basaltic crust, thick enough that the base of it transitions to eclogite, which could result in delamination or even wholesale sinking of the crust (Johnson et al., 2014; van Thienen et al., 2004), and (iii) buildup of compositional stratification if the crust and residue are able to sink into the mantle and separate at some depth: Harzburgitic residue rises to the top, while eclogitic crust sinks to the bottom, forming a depleted shallow mantle and a deep layer of subducted crust. This rapid formation of a depleted upper mantle was observed in some of the calculations of Davies (2002), Nakagawa and Tackley (2005b), Ogawa (2003), Ogawa and Nakamura (1998), but particularly in the calculation presented by Davies (2006) because of its high convective vigor (internally heated $Ra = 7.7 \times 10^{11}$) designed to represent the early Earth. Davies (2006) argued that this depletion helps plate tectonics to operate early on by preventing the formation of a thick crust; however, a crust that is thick enough to transform to eclogite can help plate tectonics because its high density can initiate subduction-like behavior (e.g., van Thienen et al., 2004).
3. Increasing chemical stratification (depending on density contrasts), possibly including a layer of MORB (i.e., subducted MORB that has undergone phase transformations to high-pressure phases) above the CMB, a depleted shallow mantle, and MORB and harzburgite trapped above and below 660 km (see **Section 7.12.3.3.4**). Increasing lateral heterogeneity as more recycled material gets subducted and stirred. It is possible that the phase changes around 660 km depth (**Section 7.12.3.3.4**) play a much larger role than they do now, because phase changes have more effect at higher convective vigor (Christensen and Yuen, 1985; Ogawa, 2003).
4. Transition to a long-term, slowly evolving equilibrium state, in which a dynamic equilibrium may be set up between segregation and entrainment and between the generation (by melting) and destruction (by mixing and/or remelting) of heterogeneity. The mantle is highly heterogeneous, containing a large fraction of recycled material. At this point, most of the mantle is depleted relative to the starting composition if the subducted crust is able to form a layer in the deep mantle (e.g., Christensen and Hofmann, 1994; Ogawa, 1997; Xie and Tackley, 2004a).
5. Possible switches in the regime can occur, such as remixing of previously formed layers (Ogawa, 1994, 1997). Other possible changes are a reduction in stratification at 660 km depth (Ogawa, 2003) or changes in the plate tectonic regime (e.g., as proposed by Davies, 1992; Sleep, 2000; Vlaar, 1986). The latter occurs to some extent in Ogawa (2007, 2013).

It is worth noting some findings of such calculations. Firstly, layers that form from segregation and settling of dense material (as discussed in **Section 7.12.3.3.3**) tend not to have sharp boundaries as do layers that are introduced a priori (**Figure 7**). This makes it more difficult for plumes to form from their tops (Tackley, 2011).

Second, changes in internal heating rate can cause changes in the regime, for example, between layered (caused by vigorous magmatism and basalt/eclogite settling) and well-mixed as occurred in the calculations of Ogawa (1997) and Ogawa and Nakamura (1998). Dupeyrat et al. (1995) also obtained a switch between layered and well-mixed modes, from a homogeneous start.

Third, shallow mantle depletion can cause some dynamic effects, as found by Dupeyrat et al. (1995) modeling upper mantle convection. A layer of depleted residue building up under the lithosphere reduces convective velocity and increases the wavelength of convective cells. When plates are included, they carry the buoyant residue down to the bottom where it can subsequently rise in the form of diapirs, which increases mixing rate (as observed in many of the previously discussed studies).

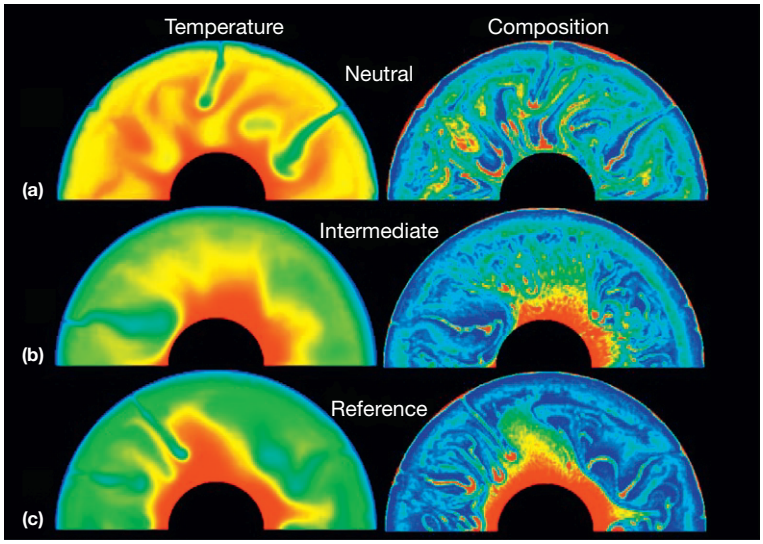


Figure 7 Laterally heterogeneous mantle and 'messy' layering resulting from continuous oceanic differentiation and recycling of MORB and residue for billions of years, from Nakagawa and Tackley (2005a). MORB, red; harzburgite, blue. All models include composition-dependent depth of the transition to perovskite (deeper for MORB) and the same compositional density contrasts in the upper mantle and top of lower mantle. The density contrast of MORB relative to pyrolite at the CMB is either (a) 0%, (b) 1.1%, or (c) 2.2%. When MORB is dense at the CMB, some of it segregates into a layer, which is intermittent in the intermediate-density case. In all cases, the transition zone is enriched and the top of the upper mantle is depleted, due to the different depths of the perovskite phase transition.

7.12.3.3.2 Magmatic heat transport

In models that include secular cooling, extrusive magmatism can be a dominant transport mechanism of heat through the lithosphere at early times. Magmatic heat transport is the sum of energy transport due to latent heat (absorption due to melting in the asthenosphere and then release when the magma solidifies as crust) and energy loss to rapid cooling of the newly formed hot crust at the surface. This 'magmatic heat pipe' mechanism is thought to be the dominant interior heat loss mechanism on Jupiter's moon Io (O'Reilly and Davies, 1981) and tends to result in a cold crust/lithosphere. Parameterized models predicted that it should have been an efficient heat loss mechanism in the early Earth (Davies, 1990). In numerical simulations for Earth by Xie and Tackley (2004b), heat transported by magma was about equally important as conductive heat loss early in the calculation and, later on, subsided to a small fraction of the total heat transport as is the case on present-day Earth. In subsequent simulations of mobile-lid convection for Earth (Nakagawa and Tackley, 2012), episodic-lid convection on Venus (Armann and Tackley, 2012), and stagnant-lid convection on Mars (Keller and Tackley, 2009; Ogawa and Yanagisawa, 2011), magmatic heat transport was found to dominate heat transport at early times. These numerical simulation results are broadly consistent with the parameterization of van Thienen et al. (2005b). This process also acts to buffer mantle temperature, giving a thermostat-like effect (Ogawa and Yanagisawa, 2011). Based on box modeling of xenon isotopes, Coltice et al. (2009) argued that the Earth's early outgassing rate was so high that magmatism was the dominant heat transport mechanism. Recently, Moore and Webb (2013) highlighted magmatic heat piping as being important in the early Earth and made a case that it is consistent with geologic observations.

These models all assume that magmatism is purely extrusive, however. Intrusive magmatism could have quite a different effect, causing a warm, soft crust rather than a cold, hard crust. As intrusive magmatism is generally thought to be more important than extrusive magmatism (Crisp, 1984), its effect on global heat transport and dynamics at early times must be studied in the future.

7.12.3.3.3 Development of a dense layer by gravitational settling of subducted crust

The idea that subducted oceanic crust might settle at the CMB was initially proposed by Hofmann and White (1982) as a way of separating the subducted crust from convective mixing long enough for it to develop the isotopic age observed in OIBs. Long-term storage of subducted oceanic crust may also help explain other geochemical signatures – for example, sulfur isotopes indicating subducted crust older than 2.45 Ga (Cabral et al., 2013). This section discusses constraints on the plausibility of this process, with a focus on MORB. It is also likely that other crustal types have been subducted although in smaller quantities than MORB, for example, Kawai et al. (2009) and Komabayashi et al. (2009) argued that significant amounts of Archean TTG and primordial anorthositic crust have been subducted, with anorthositic crust able to sink to the deep mantle and TTG either becoming trapped in the transition zone (Kawai et al., 2009) or also sinking into the deep mantle (Komabayashi et al., 2009). Tatsumi et al. (2013) found that the residue from the formation of upper continental crust (which is thought to form mainly by partial melting of arc basalts; see **Section 7.12.1.5.2**) is denser than pyrolite at all mantle depths and might therefore contribute to any chemical layering above the CMB.

7.12.3.3.3.1 Subducted MORB density contrast

A key parameter in the process of crustal settling is the density contrast of the MORB composition at the pressure and temperature range of the CMB. Unfortunately, this contrast is still not well known due to uncertainties in the properties of the mineral phases present in MORB at these pressures, although there are ongoing improvements in our knowledge of these. Historically, Ringwood (1990b) proposed that basalt remains 2–4% denser than peridotite throughout the whole mantle, with the exception of the depth range 660–720 km, where it is less dense, the consequences of which are discussed in **Section 7.12.3.3.4**. Subsequently, it was calculated that MORB might actually be less dense than pyrolite at CMB pressures (Kesson et al., 1998; Ono et al., 2001). According to Kesson et al. (1998), MORB is intrinsically 0.9% denser than pyrolite at 1100 km depth but −0.5% (28 kg m^{-3}) less dense than pyrolite at the CMB, while depleted peridotite is less dense than pyrolite at all depths: −0.8 and −0.7% at 1100 km and the CMB, respectively. At face value, this implies that MORB would be positively buoyant at the CMB, but what may matter is the density relative to residue, rather than relative to pyrolite, which is +0.2% (11 kg m^{-3}). This is because it seems likely that 95–99% of the mantle has been processed through MOR melting and thus has been differentiated into residue and MORB, with very little pyrolite remaining. In any case, 11 kg m^{-3} density difference is not enough to have a significant dynamic effect. More recently, Ono et al. (2001) calculated that although MORB is ~100 kg m^{-3} denser at the top of the lower mantle, its density profile intersects an average mantle density at around 1500–2000 km depth in the lower mantle, making MORB neutrally buoyant at ~1600 km depth and positively buoyant at greater depths.

More recent estimates of MORB density are moving back in the direction of it being denser than pyrolite at CMB pressures. From in situ x-ray observations of peridotite and NMORB and K-rich basalts at up to CMB pressures and temperatures, Ono et al. (2005b) estimated that MORB is still 40–80 kg m^{-3} denser than the average mantle at a depth of 1800 km, at which depth they had previously estimated MORB to have ~zero density anomaly (Ono et al., 2001). Guignot and Andrault (2004) refined the equation of state of phases containing Na, K, and Al and calculated MORB to be denser than pyrolite by 100–160 kg m^{-3} at the top of the lower mantle reducing to 25–95 kg m^{-3} at the CMB, which corresponds to 0.45% to 1.7% of the density at the CMB and results in a buoyancy parameter B of 0.2–0.7 (assuming a thermal expansivity of 1.0×10^{-5} K^{-1} and superadiabatic temperature contrast of 2500 K) or 0.4–1.4 (assuming a local temperature contrast of 1200 K). Hirose et al. (2005) measured physical properties of MORB samples at pressures and temperatures up to those near the CMB and found that MORB is substantially denser than PREM throughout the lower mantle and by about 200 kg m^{-3} (3.6%) near the CMB, corresponding to a buoyancy ratio B of 1.4 ($\Delta T = 2500$ K) or 3.0 ($\Delta T = 1200$ K). Most recently, Ricolleau et al. (2010) measured the phase assemblages and densities of a natural MORB sample at up to 89 GPa and then calculated the density profile of MORB for various compositions and temperature profiles. Depending on exact composition, they found that thermally equilibrated MORB is between 0.5% and 2% denser than PREM over the entire lower mantle range.

7.12.3.3.3.2 Regional modeling studies

The laboratory study of Olson and Kincaid (1991) is particularly enlightening because there are no questions of resolution, artificial diffusion, or other issues that plague numerical treatments of composition. They introduced compositionally stratified slabs with equally thick layers of 1% denser and 1% less dense material ($B \sim \pm 0.6$) and initially 3.5 orders of magnitude higher viscosity to the bottom of their tank and found that after a period of warming, the buoyant component of the slab underwent a Rayleigh–Taylor instability in which the low-density 'depleted' material rose through the high-density 'crustal' material, which then remained above the CMB in a 'spoke' pattern and subsequently got entrained over time, as will be discussed later. The scaled time from slab incidence to dense spokes was 1.8 Ga. In this experiment, which has a buoyancy ratio in the range discussed earlier, almost all of the dense 'crust' settled to the CMB, but this may be exaggerated because the 'crust' and 'residue' layers were assumed to be equally thick, whereas in reality, the crustal layer is much thinner.

This case of compositionally stratified slab segments dropping onto a hot CMB was studied using 3-D numerical models by Tackley (2011). The different buoyancies of the high-pressure versions of MORB and harzburgite give the slab a torque that tends to make the MORB side face down. If the slab lands basalt-side down, then MORB can peel off from its underside via fingering instabilities. If the slab lands harzburgite-side down, then MORB can segregate through harzburgite extrusion from the slab sides or by the Rayleigh–Taylor instability observed by Olson and Kincaid (1991). The fraction of MORB that segregates ranges from 0.5–0.7 if a dense layer is already present to 0.25–0.45 if no dense layer is initially present. Slabs spreading above the CMB tend to induce plumes at their edges; these plumes initially contain harzburgitic material, although may entrain basaltic material later on.

7.12.3.3.3.3 Numerical modeling: Prescribed plates

The earliest results on this issue (Gurnis, 1986a) found that with $B = 2$, basaltic tracers that were already separated from the residue component of the slab could aggregate into piles under upwellings for completely internally heated flows; however, with basal heating, they got rapidly swept back into the main flow and did not settle at all if accompanied by complementary depleted tracers. One way of understanding this result is in terms of the Stokes velocity of a sinking blob of crust: At high lower mantle viscosities, this velocity is extremely low, for example, 0.1 mm year^{-1} for a 5 km blob with mantle viscosity = 10^{21} Pa s. If, however, the strong temperature dependence of viscosity of mantle rocks is taken into account, then relatively low viscosities may be present in the hot lower TBL, which allow the crust to settle from the residue fast enough, and several studies that have included this have found significant crustal settling is possible. A similar process has been obtained in numerical experiments, but with a smaller fraction of the crustal material remaining at the CMB. Christensen (1989b) included temperature-dependent viscosity as well as a decrease in thermal expansivity with depth in a calculation with the continuous introduction of slabs containing crust and residue layers, taking care to scale the thickness of the compositional layers in proportion to the thermal thickness of the slab. He found that for a buoyancy ratio ~1,

between 5% and 25% of the basaltic crust settled at the CMB. Again, the observed process was a Rayleigh–Taylor instability of the residue component, leaving behind some fraction of the crust. With constant viscosity, no segregation occurred, consistent with Gurnis (1986a). Interestingly, if a layer of dense crustal material already existed, a larger fraction of the incoming basalt was able to segregate and settle. These calculations were further developed in Christensen and Hofmann (1994) in which it was found that for $B=1.5$, ~14% of the total amount of crust in the mantle was in 'pools' (similar to piles) above the CMB, implying that ~1/6 of the subducted crust initially settled into a layer at the bottom. Higher Rayleigh number decreased the size of the bottom 'pools,' but higher B increased them (e.g., 26% of the total crust for $B=2.25$). Making the viscosity less temperature-dependent reduced the amount of settling, as with previous results. Interestingly, the bottom pools were not 100% crust – rather their crustal content was ~60%. Further testing and extension of such models to higher Ra was done by Brandenburg and van Keken (2007), who confirmed that the trend of less basalt segregation with higher Ra continues to higher Ra up to 2.10^7. This approach, with the addition of force-balanced plates, was subsequently extended to 2-D cylindrical geometry (Brandenburg et al., 2008).

The first such models that had an Earth-like convective vigor were presented by Davies (2002), who assumed a density difference of up to 100 kg m^{-3} but a constant $\alpha=2\times10^{-5}\,K^{-1}$, about a factor of 2 higher than realistic at the CMB. In contrast to the models of Christensen and Hofmann (1994), the model was heated entirely from within, with no hot TBL and hence no low-viscosity region at the bottom (the fraction of basal heating was arguably higher than realistic in the models of Christensen and Hofmann (1994)). Nevertheless, despite the absence of a low-viscosity TBL, a high α, and modest compositional density difference, a significant amount of basalt settled above the CMB, although less than obtained by Christensen and Hofmann (1994). Davies (2002) suggested other mechanisms that assist crustal settling, such as thickening of the crust due to slab buckling (the Stokes velocity increases as size^2), and the fact that the vertical velocity approaches zero toward the lower boundary (making it more difficult for the convective flow to entrain sinking blobs). All these mechanisms may play a role, but given the findings of Christensen (1989b) and Christensen and Hofmann (1994) that temperature-dependent viscosity facilitates crustal settling above the CMB, it is likely that Davies (2002) would have obtained more crustal settling, if the lower boundary region were less viscous. Such calculations were extended to early-Earth conditions (i.e., higher internal temperatures) by Davies (2006, 2007). At high internal temperature, corresponding to a low internal viscosity and high effective Ra, more basalt settling was observed, with a strongly depleted region forming in the upper part of the mantle. As a result of this depleted region, the thickness of early oceanic crust was no higher, or sometimes lower, than that under modern-day conditions.

7.12.3.3.3.4 Modeling studies: Self-consistent plates

The studies discussed in this section so far have controlled the surface velocities (to make them platelike) and the rate at which compositionally distinct material is injected. A number of other studies (e.g., Nakagawa and Tackley, 2005c; Ogawa, 2000a, 2003; Xie and Tackley, 2004a) attempted to treat these aspects self-consistently, with the crustal production rate determined by melting induced when the temperature reaches the solidus and the surface velocities determined by solution of the governing equations, often using a rheology that facilitates a 'platelike' behavior (e.g., Moresi and Solomatov, 1998; Yoshida and Ogawa, 2004; see also **Chapter 7.07**). It is not clear that this makes the resulting calculations more representative of Earth, but they are different from the previously discussed studies in other ways that make it instructive to discuss the results here.

The 2-D box calculations of Ogawa (2003), building on a series of earlier papers (Kameyama et al., 1996; Ogawa, 1988, 1993, 1994, 1997, 2000a,b; Ogawa and Nakamura, 1998), include self-consistent differentiation, a plate treatment involving hysteresis (Yoshida and Ogawa, 2004), and a cooling mantle with decaying heat sources. With MORB having a density anomaly of 120 kg m^{-3} and a somewhat high constant thermal expansivity of $3\times10^{-5}\,K^{-1}$, subducted MORB settles into a layer above the CMB that accumulates with time, while subducted residue undergoes diapir-like instabilities raising it back through the mantle. The basaltic layer has large undulations and typically has a thickness of ~1/4 of the domain depth, which scales to 500 km. Subsequently, the influence of variations in internal heating rate was explored (Fujita and Ogawa, 2009; Ogawa, 2007). High internal heating rates (representative of early Earth) led to unstable basaltic piles and episodic bursts of upwelling that caused chaotic plate motions, whereas lower internal heating rates (representative of recent Earth) led to stable basaltic piles and more stable plate configurations. Such models cause a variety of plume types (Ogawa, 2010).

The 2-D cylindrical calculations of Tackley and Xie (2002), Xie and Tackley (2004a,b), and Nakagawa and Tackley (2004b, 2005a,c) include self-consistent magmatism, temperature- and depth-dependent viscosity and yielding-induced plate tectonics (but with no hysteresis), and decreasing convective vigor due to decaying heat-producing elements and a cooling core. The material properties such as thermal expansivity have reasonable depth dependencies. Various density profiles for the MORB component relative to the residue component in the lower mantle are considered, and the density of the average composition ('pyrolite') is assumed to be a weighted average of the densities of the two end-members, consistent with the small-scale structure being a heterogeneous mixture of the two end-members, but not primitive never-melted pyrolite. These calculations show that even for a relatively small density contrast for MORB in the deep mantle (e.g., 62 kg m^{-3}; Nakagawa and Tackley, 2005a), subducted MORB settles at the base and builds up into large discontinuous 'piles' of material extending up to ~700 km into the mantle. With double this density contrast (123 kg m^{-3}; Nakagawa and Tackley, 2005a; Xie and Tackley, 2004a,b), a continuous undulating layer of dense material is formed, again up to 500–700 km thick. When basalt is neutrally buoyant at CMB pressures, the CMB region is still heterogeneous because subducted crust and residue reside at the bottom for some period, while they heat up sufficiently to be stirred back in to the convective flow (Xie and Tackley, 2004a). If there is crossover in density, with MORB positively buoyant at the CMB, then a layer of depleted

residue can build up above the CMB, with blobs of MORB residing in the mid lower mantle where it is neutrally buoyant (Xie and Tackley, 2004a,b). The effect of initial CMB temperature on the present-day thermal and chemical state of the mantle and core was found to be minimal (Nakagawa and Tackley, 2010) unless the CMB became completely blanketed in MORB, because cases with different initial CMB temperatures follow convergent evolutions. Similarly, initial mantle temperature was also found to have a minimal effect on the present-day structures (Nakagawa and Tackley, 2012), even though there were substantial differences early on.

Recent 2-D Cartesian models with relatively simple physical assumptions (Li and McNamara, 2013) suggest far less basalt settling than observed in the various models discussed earlier. The reconciliation of this is not clear at present. While using a (realistically) thin crust may be one factor, other possible factors are the slabs being much broader than realistic, very strong hot plumes (due to pure basal heating in most cases; i.e., zero internal heating) and a weak dependence of viscosity on temperature. It will be important to understand this by future modeling (see also **Chapter 7.11**).

7.12.3.3.3.5 Differing basalt fractions

In the calculations of Ogawa (2003), Tackley et al. (2005), and subsequent works by the same authors, the basal layer of dense material can become much thicker than observed in the calculations of Brandenburg and van Keken (2007), Christensen and Hofmann (1994), and Davies (2002, 2006, 2007), a major reason being that the total amount of MORB that can be produced by melting of primitive material is higher. Brandenburg and van Keken (2007), Christensen and Hofmann (1994), and Davies (2002) limited the total amount of MORB in the system to 10% or 12.5%, whereas the other authors assumed, based on the fraction of garnet + pyroxene in pyrolite being 30–40%, that up to 30% or 40% basalt fraction can be generated by melting. In fact, Xu et al. (2008) calculated that the amount of MORB it is possible to extract from a pyrolytic composition is 20%, so this is probably a reasonable compromise and has been used in recent studies (e.g., Nakagawa and Tackley, 2010, 2012; Nakagawa et al., 2009, 2010, 2012). In reality, basalt is not the only composition that can be produced by partial melting – pyrolite can melt up to 100% with the bulk composition of the melt changing depending on the degree of melting (e.g., Asimow et al., 2001; McKenzie and Onions, 1992), something that could be accounted for in future models.

7.12.3.3.3.6 More or less basalt settling at higher Ra?

In the calculations of Christensen and Hofmann (1994) and Brandenburg and van Keken (2007), it was observed that the amount of basalt segregation above the CMB decreases with increasing *Ra*. Seemingly in contradiction of this are other results (e.g., Davies, 2007; Nakagawa and Tackley, 2012; Ogawa, 2007) in which the amount of basalt segregation stays similar or increases with increasing *Ra*. A possible reconciliation of these different trends is the different assumptions made about crustal thickness. In Christensen and Hofmann (1994) and Brandenburg and van Keken (2007), crustal thickness is scaled with TBL thickness, so higher *Ra* is accompanied by thinner crust. In contrast, in Davies (2007), Nakagawa and Tackley (2012), and Ogawa (2007), higher *Ra* is generally associated with a thicker crust, because melting is based on a solidus and higher *Ra* typically means a hotter mantle hence deeper melting. Thus, in the latter models, the rate of supply of MORB to the deep mantle increases greatly with *Ra*. Additionally, it seems likely that separation of crust from the slab above the CMB is easier when the crust is thicker and/or the viscosity is lower.

7.12.3.3.3.7 2-D versus 3-D

Most of the previous calculations of long-term mantle evolution are in 2-D, so an important question is whether the results would be much different in 3-D. For example, as discussed in **Section 7.12.3.2.1**, different scaling laws were found (i.e., with different exponents for *B* and *Ra*) for entrainment by 2-D sheets and 3-D plumes in the laboratory experiments of Davaille (1999b). A factor to bear in mind is that toroidal motion, which is thought to increase the mixing rate (**Section 7.12.2.3.4**), is negligible in the deep mantle. For evolution calculations covering the age of the Earth, however, Nakagawa and Tackley (2010) compared simulations computed in full 3-D spherical geometry with identical calculations computed in a 2-D spherical annulus (Hernlund and Tackley, 2008) and found essentially no difference in averaged diagnostics such as mantle and core thermal evolution. In models with imposed plate motions and the continuous introduction of dense basaltic crust by subduction, Huang and Davies (2007a) found essentially the same results in 3-D as Davies (2002) did in 2-D. In regional models of slab–CMB interaction (Tackley, 2011), it was found that 2-D models gave a good guide to the amount of basalt settling in 3-D models. For mixing of passive tracers, Coltice and Schmalzl (2006) found no difference in mixing rates between 2-D and 3-D. Of course, the detailed planform of dense material above the CMB requires 3-D geometry to establish.

7.12.3.3.4 Chemical layering induced by phase transitions

The intriguing possibility exists that one or more phase transitions near or below 660 km depth could enforce or enhance chemical layering. Much research has focused on the effect of a single phase transition – the endothermic spinel to perovskite + magnesiowüstite transition at around 660 km depth, which, according to dynamic models that assume this transition applies to all mantle material, can cause partial layering of thermal convection, with the degree of flow stratification very sensitive to the value of the Clapeyron slope (e.g., Christensen and Yuen, 1985; Machetel and Weber, 1991; Peltier and Solheim, 1992; Tackley et al., 1994; Weinstein, 1993). In reality, however, olivine makes up only ~60% of the mantle mineral assemblage, and the phase changes in the other pyroxene–garnet system occur at different pressures. In particular, in the pyroxene–garnet system, the transition to the perovskite phase occurs with a positive Clapeyron slope and probably at a greater depth in the range 720–750 km. This has two effects: (i) the effect of the combined phase transitions on purely thermal convection is much reduced relative to what would occur if the olivine transitions applied to all mantle material, and (ii) subducted oceanic crust is positively buoyant (less dense than other compositions) in a region from ~660 to 730 km depth, whereas it is negatively buoyant (denser) at other pressures (Irifune and Ringwood, 1993). This has been

suggested to cause the separation of slab components in the vicinity of 660 km, with the accumulation of 'megaliths' (Ringwood, 1991; Ringwood and Irifune, 1988).

Estimates of the depth extent over which this anomalous buoyancy occurs have varied widely: A recent study of Ono et al. (2001) puts it at 660–720 km. A possible intermediate transition to ilmenite (Chudinovskikh and Boehler, 2004) would reverse this picture and increase the layering (van den Berg et al., 2002), but it is doubtful whether ilmenite occurs in a pyrolite composition (S. Ono, personal communication).

7.12.3.3.4.1 A single endothermic phase transition

If a chemical boundary already exists at a similar depth to the 660 km phase transition, Christensen and Yuen (1984) showed that their effects on deflecting a subducting slab and therefore causing flow layering can be added linearly. For the parameters they assumed, flow layering required a 4.5% chemical contrast or a Clapeyron slope of -5 MPa K^{-1} or a linear combination of the two. If chemical stratification does not, however, already exist, can an endothermic phase transition dynamically induce it? For a long-wavelength initial chemical heterogeneity, Weinstein (1992) showed that it can for certain combinations of chemical buoyancy ratio and Clapeyron slope. The endothermic phase was shown to act as a filter, trapping chemically dense material below it and chemically less dense material above it, because the total buoyancy (thermal + chemical) of these types of material is insufficient to overcome the phase change buoyancy. For this mechanism to operate, the required chemical buoyancy contrasts were in the range 1.1–1.5% (assuming thermal expansivity $= 2 \times 10^{-5}$ K^{-1}) – much lower than those required to cause layering on their own. The required Clapeyron slope was, however, larger than realistic, in the range -5 to -8.6 MPa K^{-1}, though this is partly because the calculations were performed at a low convective vigor ($Ra = 10^5$), and would be substantially reduced if an Earth-like vigor could be reached (Christensen and Yuen, 1985). For small-scale heterogeneities introduced by the continuous injection of compositionally layered slabs, Mambole and Fleitout (2002) showed that small but measurable chemical stratification can be introduced by Clapeyron slopes as weak as -1 MPa K^{-1}, with the amount of chemical stratification increasing steadily with the magnitude of the Clapeyron slope. With only the endothermic transition active, the dense component, which is the enriched, subducted MORB, is more prevalent in the lower mantle, while depleted, relatively buoyant material is more prevalent in the upper mantle. This result was confirmed by Nakagawa and Buffett (2005) in the case where the compositional anomalies were generated by melting-induced differentiation and by van Summeren et al. (2009) for the case where compositional anomalies are the result of primordial layering. None of these studies accounted for the change to a positive Clapeyron slope at high temperatures, such as might be relevant to plumes (Hirose, 2002). If the full system of phase transitions is considered, the sense of layering is reversed, as discussed in the succeeding text.

7.12.3.3.4.2 Composition-dependent phase transitions and subducted slabs

An important question is whether the basaltic crust and harzburgitic residue can separate out due to the density inversions that occur around and below 660 km depth, as suggested by Ringwood (1994). Three studies have demonstrated that this does not occur for slabs with a simple rheology, due to the strong viscous coupling between the crust and the rest of the slab (Christensen, 1997; Gaherty and Hager, 1994; Richards and Davies, 1989), whereas van Keken et al. (1996) showed that if a weak layer separates the crust from the rest of the slab, crustal separation may be possible. While Richards and Davies (1989) investigated constant-viscosity slabs, Gaherty and Hager (1994) studied slabs with temperature-dependent viscosity falling vertically onto a viscosity increase at 660 km depth and a local density inversion below this depth, but their model did not include density anomalies due to phase change deflection or net compositional buoyancy of the slab below 660 km, which was an important component of the proposal of Ringwood and Irifune (1988). Thus, Christensen (1997) included these effects as well as a more realistic slab angle associated with trench migration at the subduction zone, with basalt buoyant between 680 and 900 km, but still found that compositional effects are negligible for 100 Myr old plates, and although they could cause younger plates to stagnate at lower trench velocity, they did not prevent eventual slab penetration into the lower mantle. The rheology of the slab is, however, more complicated than simple temperature dependence (Karato et al., 2001). van Keken et al. (1996) considered a slab with an idealized rheological sandwich containing a soft (low-viscosity) layer between the 'hard garnet' crust and the cold viscous slab interior and showed that the crust can separate once the slab passes below 660 km depth, if most of the slab is ten times more viscous than the surrounding material and the weak peridotitic layer is 10 or 100 times less viscous than the rest of the slab. This model underestimates the viscosity of the slab based on simple viscosity laws, but once the peridotitic part of the slab passes through the spinel to perovskite + magnesiowüstite transition, it may be considerably weakened due to transformational superplasticity associated with grain-size reducing by the phase transition (Karato, 1995), whereas the crustal component would not. In summary, with 'conventional' rheology, compositional effects play only a minor role in slab interaction with the 660–800 km region and separation of slab components at 660 km depth is not possible, but if there is substantial weakening of the subcrustal layer due to transformation superplasticity or other effects, then separation of the basaltic crust might occur.

7.12.3.3.4.3 Composition-dependent phase transitions in a convecting system

Regardless of the viability of slab components separating at around 660 km depth, several studies have shown that once slabs reach the bottom TBL and heat up, the compositionally distinct components can indeed separate (e.g., Christensen and Hofmann, 1994; Ogawa, 2000b; Olson and Kincaid, 1991; Tackley, 2011); thus, the question of how already-separated crust and residue components interact with the phase transitions is of great relevance. The stability of a transition layer (consisting of an inseparable mixture of MORB-derived eclogite and harzburgite) that has already formed in the 660 region to thermal convection was studied by Christensen (1988), who found that it could only survive for a few hundred million years, but this model did not allow for separate basalt and

harzburgite components nor the possibility of recharge of these components. Models in which differentiated components are constantly introduced by subduction (Fleitout et al., 2000; Ogawa, 2000a) show the development of a strong compositional gradient across the 660, with depleted harzburgitic material accumulating at the top of the lower mantle, enriched basaltic material accumulating at the bottom of the upper mantle, and a chemical gradient present throughout the entire mantle, as evident in many later calculations (e.g., Armann and Tackley, 2012; Davies, 2008; Nakagawa et al., 2009; Ogawa, 2007; Tackley and Xie, 2002; Xie and Tackley, 2004a,b). This entrapment of distinct components typically does not occur when a downwelling first encounters the region, but rather after the crust and residue components have separated in the deep mantle and are subsequently circulating as 'blobs.' That the chemical and flow stratification is due mainly to the different depths of the perovskite transition in garnet and olivine systems rather than Clapeyron slope effects was demonstrated by Tackley et al. (2005). Nakagawa and Buffett (2005) found that slab penetration into the lower mantle is accompanied by an upwelling counterflow of depleted material in the area surrounding the slab, which is a mechanism for generating depleted regions in the upper mantle.

Multicomponent phase transitions are not the only mechanism that can cause an accumulation of subducted basalt above 660 km: Davies (2002, 2006) showed that a low-viscosity (high Rayleigh number) upper mantle in combination with a viscosity jump at 660 km depth can also cause this.

In summary, it is quite likely that local chemical stratification exists across the 660 km interface caused by differing depths of the perovskite transition in the olivine and pyroxene–garnet systems. There is, however, still uncertainty as to these depths. This local chemical stratification around 660–720 km is a prediction that could be tested seismically, which would help to determine whether the currently estimated depths of these phase transitions are indeed correct. For additional information on phase transitions, see **Chapter 2.08**.

7.12.3.3.4.4 Postperovskite phase transition

A factor that decreases the stability of a deep, dense layer is the recently discovered postperovskite phase transition (Murakami et al., 2004; Oganov and Ono, 2004). The positive Clapeyron slope of this transition acts to slightly destabilize the lower TBL (Nakagawa and Tackley, 2004a) and any resulting chemical layering (Nakagawa and Tackley, 2005b; Tackley et al., 2007). The destabilization effect is however relatively small and seems to be lower than the uncertainties in MORB density discussed earlier and uncertainties due to composition dependence of the PPV transition (e.g., Murakami et al., 2005; Ono et al., 2005a). In any case, because dense material tends to accumulate under hot upwellings, and the positive Clapeyron slope leads to an anticorrelation between the phase transition depth and temperature, the occurrence of a thick PPV layer is anticorrelated with the occurrence of thick pile of dense material (e.g., Nakagawa and Tackley, 2006). If the CMB is in the postperovskite stability field, then a continuous but varying-thickness layer of postperovskite is expected. However, because of the rapid increase of temperature with depth in the lower TBL, it is possible that the CMB is in the perovskite stability field, which would result in a double crossing of the perovskite to postperovskite phase boundary, and postperovskite not occurring at all in hot regions (Hernlund et al., 2005; see also **Chapter 7.11**).

7.12.3.3.4.5 Self-consistent mineralogical treatments

Instead of parameterizing each individual phase transition, another approach is to calculate the stable mineral assemblage as a function of pressure, temperature, and composition using free-energy minimization and then calculate the resulting density and other physical properties (e.g., Connolly, 2005; Xu et al., 2008). Nakagawa et al. (2012) found that the resulting dynamics and structures were similar to those using a simpler parameterization provided that the latter takes both olivine and nonolivine components into account. Such an approach is a useful framework for evaluating the effect of variations in composition (the exact MORB composition was found to be important by Nakagawa et al. (2009)) and for naturally generating seismic velocity structures for comparison with seismological models (Nakagawa et al., 2010).

7.12.3.3.5 The effect of chemical layering on core heat flow and planetary thermal evolution

As was noted early on (e.g., Christensen and Hofmann, 1994), the presence of a compositionally dense layer above the CMB reduces the heat flow out of the core. This reduction can be dramatic if the dense layer completely covers the core and should be taken into account in thermal evolution calculations (e.g., McNamara and Van Keken, 2000). Unfortunately, when the layer undulates greatly and is discontinuous, such that some patches of the core are not covered, the appropriate parameterization of core heat flux is unknown, so numerical simulation becomes the preferred approach.

A long-standing paradox in modeling Earth's thermal evolution is explaining how the heat flux out of the core can stay large enough over billions of years to maintain the geodynamo, without growing the inner core to a much larger size than observed. Parameterized models of core–mantle evolution in which no complexities are present predict a final inner core size that is much larger than that observed (e.g., Labrosse, 2003; Labrosse et al., 1997; Nimmo et al., 2004). As the heat conducted out of the core is determined by what is happening in the mantle, the time history of CMB heat flow is an important quantity to investigate using numerical models, both from the point of view of understanding core evolution and in providing an additional constraint on mantle models.

Nakagawa and Tackley (2004b) used a thermochemical convection model similar to those already presented but with a parameterized core heat balance based on Buffett et al. (1992, 1996) to show that the presence of a layer of dense material above the CMB, either primordial or arising from segregated crust, substantially reduces the CMB heat flow, hence inner core growth, something also shown by models in which the dense layer is parameterized (Costin and Butler, 2006). A problem is CMB heat flow can be reduced so much that it is insufficient to facilitate dynamo action and can even become zero or negative. Because of this, a discontinuous layer, rather than a global layer, was found to be the most promising scenario. Nakagawa and Tackley (2005a) subsequently found that it is difficult to match the constraints of correct inner core

size and enough heat flow out of the core, unless there is radioactive potassium in the core (e.g., Nimmo et al., 2004); in that case, viable evolution solutions were obtained, with a trade-off found between crustal density and core K content. Subsequently, it was found that initial core temperature (Nakagawa and Tackley, 2010) and initial mantle temperature (Nakagawa and Tackley, 2012) make little difference to the present-day core and mantle state.

The recent finding that the thermal conductivity of the core is two to three times higher than previously thought (de Koker et al., 2012; Gomi et al., 2013; Pozzo et al., 2012) makes it more difficult to explain core thermal evolution because a much higher core heat flow (at least ~12 TW) is needed for a geodynamo to exist. The higher core conductivity does not influence mantle dynamics, but it does severely restrict the range of parameters for which a successful core evolution can be obtained (Nakagawa and Tackley, 2013). Having a primordial layer in addition to subsequent basaltic segregation may be necessary to obtain a successful evolution (Nakagawa and Tackley, 2014).

As was discussed in an earlier section, chemical layering has a strong influence on plume temperature and dynamics. A systematic study of the effect of chemical layering on plume heat flux, plume excess temperature, and upper mantle temperature was conducted by Zhong (2006) in 3-D spherical geometry, assuming a sharp chemical interface and a global layer and moderately temperature-dependent viscosity. The possible range of parameters (layer depth and internal heating rates) was constrained by comparing to observations. The preferred model is one in which the dense layer is relatively thin (<350 km) and the internal heating rate in the upper layer is approximately three times higher than that inferred for the MORB source region, which, curiously, is similar to the heating deficit inherent in the 'heat–helium paradox' (Ballentine et al., 2002).

7.12.3.4 On the accuracy of numerical thermochemical convection calculations

A fundamental problem with treating chemical variations numerically is the negligible diffusivity of chemical variations, which leads to the possibility of sharp interfaces and very narrow lamella (e.g., less than one grid spacing), both of which are very difficult to treat numerically. Two major classes of method are used in the mantle dynamics community to track compositional variations: field-based approaches on a fixed (Eulerian) grid and Lagrangian markers/tracers/particles that are advected through the grid. Various benchmark tests of these have been performed in the mantle convection community (Tackley and King, 2003; van Keken et al., 1997). See **Chapter 7.05** for further discussion.

Field-based approaches have the advantage of computational speed and little additional computational complexity but the disadvantages of numerical diffusion, which smears out sharp features, and numerical dispersion, which causes artifacts like overshoots and ripples. Other research communities that employ computational fluid dynamics techniques (e.g., the atmospheric science community) face similar problems, and there is much ongoing research into improving advection methods, a small sampling of which is Muller (1992), Smolarkiewicz and Margolin (1998), Xu-Dong and Osher (1996), and Yabe et al. (2002). In the mantle dynamics community, the relatively straightforward interface-sharpening scheme of Lenardic and Kaula (1993) can yield a dramatic improvement in the advection of interfaces (e.g., Tackley and King, 2003).

Tracer-based approaches have the advantage of minimal numerical diffusion (a small amount does exist because of finite error in updating each tracer's position), thus the apparent ability to represent discontinuities and sub grid-scale features, but the disadvantage that one needs 5–50 times as many tracers as grid points to prevent statistical noise from unduly influencing the solution (e.g., Christensen and Hofmann, 1994; Tackley and King, 2003), which adds a considerable computational burden. This noise arises because in order to calculate the buoyancy caused by compositional variations, some averaging must be done over tracers in the vicinity of each grid or nodal point. It has been found that taking a local average of the C value of each tracer (as exemplified by the 'ratio method') is far less noisy than methods that rely on a number count of tracers (the so-called absolute method), regardless of the underlying numerical scheme (Tackley and King, 2003). Another type of tracer method tracks only the interface between two compositions and is thus far more computationally efficient than methods where tracers exist everywhere (e.g., Lin and Van Keken, 2006a; Schmalzl and Loddoch, 2003), though it is not suited for continuously varying chemical fields and the computational needs increase up to exponentially with time as the interface becomes more complex, unless interface simplification algorithms are applied (Schmalzl and Loddoch, 2003).

Benchmarks of these various methods for cases with an active compositional layer (Tackley and King, 2003; van Keken et al., 1997) indicate a sensitivity of the results to the details of the method being used, including a strong sensitivity of the entrainment rate of a stable compositional layer to resolution (both grid and number of tracers) (Tackley and King, 2003). This means that the various numerical results reviewed in this section may not be quantitatively accurate in the amount of entrainment that exists as a function of density contrast except in calculations where the grid is greatly refined in the region of entrainment (e.g., Lin and Van Keken, 2006a; Zhong and Hager, 2003). However, convergence tests suggest that they are qualitatively correct in the general planforms and features that arise. This is an issue that will require more attention in the future, particularly as more calculations are performed in 3-D.

7.12.4 Convection Models Tracking Trace-Element Evolution

7.12.4.1 Noble Gases

7.12.4.1.1 Helium and argon

The evolution of noble gases in a mantle that includes outgassing and radiogenic ingrowth but not major-element differentiation was studied by van Keken and Ballentine (1998, 1999). These studies investigated various proposed mechanisms for maintaining a relatively undegassed lower mantle over billions of years, finding that neither high deep mantle

viscosity, a strongly endothermic phase transition at 660 km depth, nor temperature-dependent viscosity is capable of causing a relatively undegassed deep mantle, in contrast to earlier modeling at lower convective vigor (Gurnis and Davies, 1986a). They did find that the observed total amount of ^{40}Ar outgassing is consistent with whole-mantle convection over geologic history, although the convective vigor in their models was constant with time. In that case, 50% degassing of radiogenic ^{40}Ar implies ~70% degassing of primitive isotopes.

van Keken et al. (2001) extended the analysis to the heat–helium imbalance (**Section 7.12.1.6.1**), tracking the relationship between the rate of helium outgassing and the heat flux. Large fluctuations in the helium outgassing were obtained, but only the most extreme fluctuations would explain the 'heat–helium' issue as it is currently understood, that is, with a ^{4}He concentration 3.5 times higher than current estimates (Ballentine et al., 2002).

The helium ratio evolution caused by subduction of differentiated oceanic plates was studied by Ferrachat and Ricard (2001), who tracked the evolution of oceanic crust and residue and showed that if oceanic crust segregates at the CMB, a large region of recycled oceanic lithosphere with high ^{3}He/^{4}He can form above it. However, in their model, compositional variations were also purely passive, that is, not influencing the flow, whereas models that account for the density anomalies of differentiated components (discussed later) have found that the depleted material, which in this case carries the high ^{3}He/^{4}He signature, tends to rise.

7.12.4.1.2 Origin of ^{3}He/^{4}He

The origin of the high ^{3}He/^{4}He end-member is discussed in **Section 7.12.1.6.1**. The two most commonly hypothesized origins have both been the foci of numerical investigations: Samuel and Farnetani (2003) and Deschamps et al. (2011) demonstrated the dynamic viability of the 'primitive' explanation, while Ferrachat and Ricard (2001) and Xie and Tackley (2004a) demonstrated the viability of the 'recycled' explanation. In any case, it is important to recall that the ratio of ^{3}He/^{4}He constantly decreases with time due to radiogenic ingrowth of ^{4}He from U and Th decay. Elemental fractionation between He and (U+Th) upon melting and outgassing of He to the atmosphere influence He/(U+Th) in the crust and residue, which over time can then develop ^{3}He/^{4}He different from primitive material. Thus, the present-day distribution of He isotope ratios reflects a combination of ingrowth, fractionation, and outgassing.

7.12.4.1.2.1 Primitive helium

Samuel and Farnetani (2003) used a fully dynamic convection model (e.g., with no imposed surface velocities) to demonstrate that a thick basal layer of primitive material, consistent with the proposal of Kellogg et al. (1999) and van der Hilst and Karason (1999), is capable of generating ^{3}He/^{4}He histograms similar to those observed for MORB and OIB. Isotopes of U, Th, K, and He were tracked in three different components: primitive, subducted MORB, and subducted residue, with the primitive layer being the 'mass balance' reservoir of U, Th, and K as well as the high ^{3}He/^{4}He component. Partitioning of trace elements between crust and residue was based on an assumed melt fraction. With a compositional density contrast of 2.4% (100 kg m^{-3}), depth-dependent thermal expansivity, temperature-dependent viscosity, and a primitive layer making up 25% of the mantle volume, it was found that ^{3}He/^{4}He values in MORB cluster in the observed range, whereas OIB samples, which were associated with plumes forming from the layer interface, display a much wider spread because they contain a heterogeneous mixture of the three different components. Heterogeneity dispersed more rapidly in the upper layer than in the lower layer due to the higher convective vigor there. Lower layer material was entrained into the upper layer much more rapidly than upper layer into the lower layer: after 2.1 Gy, the percentages were 25% and 3%, respectively. Deschamps et al. (2011) performed 3-D simulations of mantle convection with a relatively small primitive dense layer giving a good match to probabilistic seismic tomography (Deschamps and Tackley, 2009) and showed that the amount of entrainment into plumes is in the range required to give the observed helium ratios of OIBs; therefore, the models are consistent with the seismological and geochemical constraints considered.

7.12.4.1.2.2 Recycled helium

As discussed in **Section 7.12.4.1.1**, Ferrachat and Ricard (2001) found that a high ^{3}He/^{4}He repository of subducted residue built up in the deep mantle in a steady-state flow with no compositional buoyancy. Studies that do include compositional buoyancy find, however, that the depleted residue tends to rise to the shallow mantle, as in the preliminary calculations of Tackley and Xie (2002), in which high ^{3}He/^{4}He residue aggregates near the top due to its buoyancy, where it is sampled by mid-ocean ridge volcanism. Thus, Xie and Tackley (2004a) investigated the possibility that active (buoyant) compositional variations caused by crustal production may play an important role in the evolution of noble gas isotopes and the generation and maintenance of distinct end-members, using numerical convection models with a self-consistent melting criterion and platelike dynamics. The distribution of ^{3}He/^{4}He ratios in the modeled MORB source region was found to be highly dependent on the density contrasts between the different components (primitive, residue, and basalt) and compatibility of He relative to U and Th. While in most cases the ^{3}He/^{4}He was scattered to too-high values, two parameter combinations were identified that led to a 'realistic' MORB-like distribution of ^{3}He/^{4}He (i.e., clustered around 8–10 times atmospheric): (i) He less compatible than U, Th, and crust dense in the deep mantle; in this case, the shallow mantle contained a large proportion of depleted residue, which had MORB-like ^{3}He/^{4}He, and the high ^{3}He/^{4}He component was primitive material. (ii) He equally compatible as U, Th, and crust buoyant in the deep mantle; in this case, subducted crust in the shallow mantle brought the ^{3}He/^{4}He into the correct range; both primitive material and residue had the same 'primitive' ^{3}He/^{4}He.

These calculations demonstrate that a MORB-like shallow mantle ^{3}He/^{4}He and the existence of a high ^{3}He/^{4}He end-member are consistent with whole-mantle convection from a homogeneous start but with the ambiguity of two alternative explanations. Indeed, there may be a continuum of 'successful' solutions in (partition coefficient and buoyancy) space, which would require the consideration of additional constraints to constrain further. These calculations did not attempt to

distinguish between MORB and OIB melting signatures. Thus, more investigations are necessary, preferably in 3-D.

There has also been progress on understanding helium ratios using conceptual or box models discussed later (Section 7.12.5.1).

7.12.4.2 Other Isotopes (U, Pb, Nd, and Sm) and the Origin of HIMU

The hypothesis that segregation and sequestering of subducted MORB is responsible for the HIMU component and related Pb–Pb isochron (Hofmann and White (1982)) was investigated in pioneering calculations by Christensen and Hofmann (1994) using a numerical mantle convection model that included tracking of crust residue differentiation and isotopes of trace elements U, Pb, Sm, and Nd, which partitioned between the crust and residue on melting. The model was partly heated from within and partly from below and had a kinematic time-dependent platelike upper boundary condition. As discussed earlier, a fraction of the subducted crust segregated into a layer above the CMB, and with some parameter combinations, realistic Pb–Pb isotope diagrams and 'ages' were obtained. The isotope age depended on the length of time subducted MORB resided near the CMB. However, the authors found it necessary to run the calculations for only 3.6 Gy instead of the full age of the Earth; otherwise, the isotopic age was too high, contrary to previous expectations from mixing time estimates (reviewed in Section 7.12.2.6.4) that it would be difficult to obtain a large enough age. This would be appropriate if the mantle were completely mixed 3.6 Gy before present. Furthermore, their model mantle had a constant convective vigor (e.g., velocities and rate of melting), rather than, as discussed earlier, a vigor that decreases with time as the mantle temperature and radiogenic heat production decrease.

The calculations of Xie and Tackley (2004b) followed a similar approach to those of Christensen and Hofmann (1994), but with a convective vigor that decreases with time due to secular cooling and the decay of heat-producing elements, and the investigation of 4.5 Gy run times as well as 3.6 Gy. Technical differences were the use of a viscoplastic rheology to allow platelike behavior of the lithosphere (Moresi and Solomatov, 1998), rather than kinematic velocity boundary conditions, and a more self-consistent treatment of melting and crustal generation based on comparing the temperature to a solidus. The higher convective vigor early in the calculations plus the longer evolution resulted in a lot of differentiation early in the modeled evolution, which resulted in Pb–Pb ages that are much too large (3.39 Gy for the reference case). This large age is not surprising, because the rate of melting is very high early in the evolution, so early differentiating material dominates the isotopic signature. Additionally, early differentiated material has developed a more radiogenic signature so influences the least squares fit more than recently differentiated material. This finding is consistent with previous attempts at modeling lead isotope ratios starting from 4.5 Gy ago that obtained unrealistically high slopes in $^{207}Pb/^{204}Pb$–$^{206}Pb/^{204}Pb$ space (Allegre et al., 1980; Armstrong and Hein, 1973). It is also consistent with the numerical models of Davies (2002), in which although Pb–Pb ages were not calculated, they found mean ages (since melting) of ~2.7 Gy for MORB. Christensen and Hofmann (1994) attributed this 'too old' problem to the rapid decay of ^{235}U into ^{207}Pb in the early stage of the model, also finding that the early differentiation has a strong effect on the isotopic signatures at the end of the run. Thus, they chose to start their models 3.6 Gy ago with uniform Pb composition, which assumes that any earlier differentiation in the mantle has been completely homogenized by the convection.

Several possible explanations for this 'too old' discrepancy between models and data were investigated by Xie and Tackley (2004b). These are that HIMU material may not have been subducted into the mantle prior to 2.0–2.5 Gy BP, that melting is not rapid enough in the model (i.e., too low processing rate meaning too high residence time), that stretching of heterogeneities is inadequately represented, that the 'sampling volume' over which isotopic ratios are calculated is incorrect, that the deep mantle density of crustal material is incorrect, and that the Pb–Pb slope is made artificially large by noise associated with tracer discretization. The most successful and straightforward explanation was found to be the first (i.e., that HIMU was not subducted in the late Archean), which is discussed in more detail in the succeeding text, but it is worth briefly discussing the others. Regarding the melting-induced processing rate, this was found to be within the range of possibilities for the Earth (see discussion in Section 7.12.1.5.3). A parameterized box model indicated that only the most extreme assumptions would lead to a low enough residence time. Inadequate treatment of stretching highlights a problem with tracer treatments: whereas in reality a tracer should get exponentially stretched into a long tendril that extends over an area much wider than the sampling volume, thereby reducing its contribution to the isotopic signature of a melt, in fact, it remains as a localized entity. Xie and Tackley (2004b) tested one approximation of this: tracking the integrated strain as in Kellogg and Turcotte (1990) and McKenzie (1979) and then ignoring high-strain tracers when calculating the isotopic signature of a sampling volume. This was found to reduce the Pb–Pb somewhat, but not enough to match observations. Changing the sampling volume over which heterogeneities are averaged does not change the Pb–Pb slope, only the amount of scatter.

One possible reason for the ~1.8 Ga Pb–Pb age is that HIMU material did not enter the mantle before 2.0–2.5 Gy ago, for which there are two possible explanations. One of these is that higher subduction zone temperatures in the past may have caused essentially all Pb, U, and other trace elements to be stripped from the subducted crust by melting (Martin, 1986). Supporting this explanation are eclogite xenoliths from the upper mantle under Western Africa that are highly depleted in incompatible elements and have been interpreted as the melting residue of Archean subducted crust (Barth et al., 2001). Another possible explanation is that the HIMU endmember was not produced prior to the rise in atmospheric oxygen about 2.2 Gy ago. In this latter scenario, HIMU is produced by addition of continental U to oceanic crust by hydrothermal circulation at ridges (Elliott et al., 1999; Michard and Albarède, 1985), with efficient stripping of U from continents being accomplished by an oxygen-rich atmosphere oxidizing U to its 6+ state (Holland, 1984, 1994). This is not the only process that has been proposed to

produce HIMU: the other is preferential stripping of Pb (relative to U) from the subducted crust to the overlying mantle in subduction zones (Chauvel et al., 1995; Hofmann, 1988; Newsom et al., 1986). It is possible that both of these processes may operate. In the numerical model of Xie and Tackley (2004b), as in that of Christensen and Hofmann (1994), HIMU was produced by artificially changing the partition coefficients of U and Pb so that they fractionate on melting. In order to evaluate the effect of it not being produced prior to some point in the past, Xie and Tackley (2004b) ran cases in which the U and Pb partition coefficients are set equal for the first part of the model run and then changed to their default values at a specified point in the past. Setting the transition to 2.5 Gy before present caused a Pb–Pb isotopic age of 1.75 Gy, whereas setting it to 2.0 Gy before present caused an isotopic age of 1.4 Gy. Thus, from a modeling perspective, a lack of HIMU production prior to 2–2.5 Gy ago is an appealing mechanism for explaining the effective age of the Pb–Pb isotope diagrams, although it needs to be shown that this fits other isotope systems as well.

A mathematical analysis of the relationship between the probability distribution of heterogeneity age (i.e., time since last melting) and that of the pseudoisochron age was conducted by Rudge (2006), superseding an earlier box model approach by Allegre and Lewin (1995). He obtained simple relationships between mean remelting time and isochron age for a constant melting rate, which closely fit the Pb–Pb pseudoisochron ages in the numerical simulations of Christensen and Hofmann (1994). His more general theory was also able to fit the isochron ages of the time-dependent melting histories in the simulations and calculations of Xie and Tackley (2004b). An important finding is that for constant melting rate, the Pb–Pb pseudoisochron age is substantially higher than the mean time before remelting – in particular fitting a 2.0 Gy Pb–Pb pseudoisochron age with a constant melting rate over the Earth's history requires a time interval of only 0.5 Gy before material remelts, much shorter than seems reasonable (**Section 7.12.1.5.3**).

7.12.4.3 Extension to Include Continental Differentiation

This style of modeling was further developed by Brandenburg et al. (2008) who added Rb–Sr and Re–Os isotope systems to the U–Th–Pb and Sm–Nd included the previously mentioned studies. An additional feature of their model was the extraction of trace elements to the continental crust, which was parameterized in what the authors described as an ad hoc process. Extraction of continental crust was found necessary in order to obtain reasonable fits to data in all isotope systems. It was also necessary to change the continental crust extraction efficiency for U and Pb at 2.25 Gy before present to match Pb–Pb systematics, similar to as discussed in the previous paragraph. Interestingly, they found that ancient oceanic crust (subducted long in the past) adds an EM-I-like component, while more recently subducted oceanic crust adds a HIMU component.

'Box' modeling of the distribution of isotope ratios by sampling a heterogeneous system was performed by Kellogg (2004) and Kellogg et al. (2002, 2007) using a new type of 'box model' that allows the distribution of isotope ratios to be calculated in addition to the mean values (Allegre and Lewin, 1995). The method was used to successfully match Rb–Sr and Sm–Nd isotope ratio diagrams by differentiation and recycling in the upper mantle–continental crust system (i.e., no oceanic recycling) and subsequently extended to include the U–Th–Pb system in Kellogg (2004) and Kellogg et al. (2007). In the latter model, a lower continental crust was added but the lower mantle remained primitive. It was found that matching lead isotopes required considering recycled MORB and again a change regarding relative U and Pb behavior at 2.3 Gy before present.

From the earlier two studies, it is clear that both differentiation associated with oceanic crust production and differentiation associated with continental crust production must be considered in order for all isotope ratios to be consistent with observed ones.

The convection models of Walzer and Hendel (1997a,b, 1999, 2008) included only the melting-induced formation of the continental crust, without treating the massive differentiation caused by oceanic crust formation. Trace-element ratios of crust and residue were constrained to be those in the geochemical literature (e.g., Hofmann, 1988) rather than calculated based on partition coefficients. While the models display some interesting dynamic characteristics, they are geochemically predetermined to match observations and thus cannot provide predictions about the origin of geochemical heterogeneities.

7.12.4.4 Isotopic Variations at Hotspots and Transport by Plumes

Spatial and temporal isotopic variations within a hot-spot chain were not addressed in the previously mentioned studies, but are a notable feature of hot-spot volcanism (White, 2010); for example, the Hawaiian chain displays two isotopically distinct tracks of volcanoes often referred to as the 'Loa' and 'Kea' chains. A key question is whether this difference is caused by isotopic heterogeneity in the region feeding the presumed plume above the CMB or caused by processes in the region where the plume interacts with the lithosphere (see **Chapter 7.10**).

Investigating the feasibility of the plume source region explanation, Farnetani and Hofmann (2010) modeled in 3-D the propagation of heterogeneities from the lower TBL above the CMB up the plume conduit to the melting region beneath Hawaii, building on an earlier, 2-D axisymmetric study (Farnetani and Hofmann, 2009; see also **Section 7.12.1.4**). Blobs in the source region become filaments as they are advected up the plume. A straightforward mapping was found between isotopic heterogeneity in the shallow region and in the source region. Filaments of 10 km radius cause isotopic changes on timescale 400 ky. Subsequently, Farnetani et al. (2012) showed how a large-scale gradient in Pb isotopes above the CMB corresponding to the DUPAL anomaly maps to shallow structure. According to these modeling results, it is possible to directly link geochemical maps of OIBS to the structure and composition of the lowermost mantle.

Helium ratios display a different variation, in that the highest $^{3}He/^{4}He$ signature is observed at the most recently formed volcano, Lō'ihi. Hofmann et al. (2011) argued that this is due to processes occurring within the plume conduit at transition

zone depths. Specifically, they proposed that carbonatite melting in the transition zone followed by rapid ascent of the melt relative to the tilted plume displaces the higher $^{3}He/^{4}He$ signature toward the leading edge of the plume.

Investigating the 'shallow processes' explanation, Bianco et al. (2008) modeled plume–lithosphere interaction and melting of a plume with two geochemical components, 10% enriched (EC) and 90% depleted (DC), and found that two geochemically distinct trends (reminiscent of the 'Loa' and 'Kea' trends) could be generated by different amounts of melting, with no chemical zoning of the plume being necessary. This was further developed to include the effects of preexisting small-scale convection by Ballmer et al. (2011), who found that the interaction of the plume with small-scale convection-related washboard topography on the base of the lithosphere caused asymmetrical spreading and melting of the plume and hence in geochemistry of the erupted lavas, as well as temporal variations in the flux or erupted material (see **Chapter 7.10** for further discussion).

7.12.5 New Concepts and Future Outlook

7.12.5.1 Revisiting Noble Gas Constraints

7.12.5.1.1 Ar and He budget

Noble gas constraints were reviewed in **Section 7.12.1.6.1**. As discussed in that section, the '50% ^{40}Ar degassing' constraint is straightforward to match with whole-mantle convection, even with substantially higher processing rates in the past. A much more troubling constraint from the point of view of whole-mantle convection models is that the argon concentration in the MORB source is much lower than it would be if the Ar thought to be in the mantle were evenly distributed, which has been taken as evidence of layered convection with the lower layer containing the 'missing' argon. Along similar lines is the 'heat–helium imbalance,' that is, the concentration of He in the MORB source region seems to be too low relative to concentrations of U and Th. Recently, Ballentine et al. (2002) noted that both of these constraints are based on fluxes of ^{3}He through the oceans that represent an average over 1000 years, and if the true long-term average were a factor of 3.5 higher, both constraints would be removed. He argues that two other indicators of ^{3}He concentration in the MORB source – the concentration in the 'popping rock' and the estimate of carbon concentration in MORBs based on graphite–melt equilibrium – give estimates of ^{3}He concentration up to several times higher than obtained from the He flux through the oceans. On the other hand, Saal et al. (2002) obtained rather lower ^{3}He concentrations by considering carbon concentrations in undegassed melt inclusions. Numerical models of convection that track the relevant isotopes (van Keken et al., 2001) obtain considerable variation in the outgassing rate over timescales much longer than 1000 years.

The 'standard' reconciliation of the heat–helium imbalance invokes the concept that He and heat have different transport mechanisms, that is, heat is transported across a TBL at 660 km depth, whereas He is not. Castro et al. (2005) found that a similar concept occurs to He transport in continental aquifers, which, when applied to the oceans, suggests that the He flux into the ocean may be much lower than the flux out of the mantle, due to the action of seawater circulating through the oceanic crust (see Albarède (2005a) for further discussion).

The 'missing ^{40}Ar' constraint would also be eased if the total amount of ^{40}K in the mantle were lower than commonly assumed, which is based on BSE abundances (e.g., McDonough and Sun, 1995) combined with the K/U ratio observed in MORB (Jochum et al., 1983). It has been shown that this difficulty could be resolved if the amount of ^{40}K in the mantle were a factor ~2 than commonly estimated (Albarède, 1998; Davies, 1999), which would be the case if the K/U ratio were lower than that observed in MORB, perhaps due to fractionation of K from U (Albarède, 1998) or because the Earth is nonchondritic (Campbell and O'Neill, 2012; Lyubetskaya and Korenaga, 2007b). Lassiter (2004) suggested that K/U could be significantly lower than normally assumed, if the K depletion of oceanic crust in subduction zones is taken into account. Then, subducted MORB in the mantle would have a relatively low K/U, lowering the mantle average.

As a result of these various arguments, it seems plausible that the 'missing ^{40}Ar' and 'heat–helium imbalance' constraints are less robust than previously thought.

7.12.5.1.2 High $^{3}He/^{4}He$

Although the high $^{3}He/^{4}He$ end-member (FOZO/C) is often assumed to be primitive and undegassed, several authors (e.g., Albarède, 1998; Anderson, 1998; Coltice et al., 2000b; Coltice and Ricard, 1999, 2002; Ferrachat and Ricard, 2001) have suggested that it could instead be generated by recycling, as was discussed in detail in **Section 7.12.1.6.1** and investigated by the numerical models of Ferrachat and Ricard (2001) and Xie and Tackley (2004a) discussed in **Section 7.12.4.1**. If high $^{3}He/^{4}He$ is caused by low [U + Th] rather than high [^{3}He], then an inverse correlation is expected between [U + Th] and $^{3}He/^{4}He$, and such a correlation was found in the data plotted by Coltice and Ricard (1999). Meibom et al. (2003) pointed out that the unfiltered $^{3}He/^{4}He$ dataset for MORB has a Gaussian distribution similar to that for $^{187}Os/^{188}Os$ and proposed that both are due to mixing of radiogenic and nonradiogenic components (associated with basalt and residue respectively) with varying amounts of partial melting. They proposed that residue can retain substantial He due to the capture of He-rich bubbles in magma chambers. Further reinforcing this interpretation is a recent compilation of OIB isotope data, which shows that high ^{3}He source has the same isotopic and major-element composition as depleted mantle and therefore represents already-melted material (Class and Goldstein, 2005). Hauri et al. (1994) had earlier concluded that the FOZO represented differentiated material. Class and Goldstein (2005) found an inverse correlation between $^{3}He/^{4}He$ and Th abundance consistent with FOZO being a mixture of a small amount (~2%) primitive material plus ingrown ^{4}He. Based on the difference between plumes and OIBs, they inferred that the amount of time the high $^{3}He/^{4}He$ material has remained separate from the rest of the mantle is ~1–2 Gy. Heber et al. (2007) measured lower partition coefficients for noble gases in olivine and clinopyroxene than previously measured; however, they still concluded that residue from previously melted material could account for the high $^{3}He/^{4}He$ material.

The recycled-helium explanation is not universally accepted, however. Gonnermann and Mukhopadhyay (2007)

presented a new model for nonequilibrium degassing of CO_2, H_2O, and noble gases that explains the helium concentration paradox (i.e., low [^{3}He] in high ^{3}He/^{4}He samples). By this mechanism, the OIB source can have a large, primitive He concentration if it has more CO_2, while the MORB source has less CO_2. In the context of mantle convection, Gonnermann and Mukhopadhyay (2009) showed using a 'box' model that the lower mantle can retain relatively high ^{3}He/^{4}He despite convection and processing, including slab penetration, provided the flux of material into the lower mantle is low, at 0.5–1 lower mantle masses over the Earth's history. However, from a dynamic perspective, this amount must be regarded as impossibly low, as it is not only slabs that penetrate but also the general large-scale circulation.

Another recent proposal is that the high ^{3}He/^{4}He endmember is carried by refractory rocks, as a result of He and Ne diffusing into such rocks early in the Earth's history (Albarede, 2008). Such refractory material would be less sampled by melting and possibly less stretched and mixed.

7.12.5.2 Improved Recipes for Marble Cakes and Plum Puddings

7.12.5.2.1 Two recipes

7.12.5.2.1.1 Heterogeneous assemblage

Extreme compositional heterogeneity of the mantle, building on the 'marble cake' proposal of Allegre and Turcotte (1986), is a common theme in recent attempts to explain mantle geochemistry and contrasts with earlier views of a relatively homogeneous upper mantle, with the necessary OIB heterogeneity supplied by plumes from the deep mantle. The proposed mantle assemblage typically consists of a mostly depleted residue with approximately few % primitive material, ~10–20% recycled MORB, and approximately few % recycled sediment (derived from continental crust) and possibly continental lithosphere and/or OIB.

7.12.5.2.1.2 Two-assemblage recipe

One possible recipe is that the assemblage in the MORB source region is different from that in the plume (hence OIB) source region, but both assemblages are derived mostly by recycling processes. This is consistent with the proposal of Christensen and Hofmann (1994) and Hofmann and White (1982) that HIMU and perhaps EM components arise from subducted crust that has segregated into a layer above the CMB. This recycled deep layer can also be the location for storing incompatible elements that are 'missing' according to mass budget calculations (Coltice and Ricard, 1999) and have higher concentrations in MORB than in primitive material, requiring a smaller storage area. The high ^{3}He/^{4}He component might be caused by strips of former oceanic lithosphere (Coltice and Ricard, 1999, 2002) as discussed in the previous section. In the marble cake of Coltice and Ricard (1999, 2002), this subducted lithosphere forms a layer above D″, accounting for different MORB and OIB signatures, but it is not clear that such a layer would be stable. A deep layer of residue was obtained in the numerical calculations of Ferrachat and Ricard (2001), but compositional density contrasts were not accounted for in those calculations, and when they are, residue tends to rise into the upper part of the mantle (e.g., Ogawa, 2003; Tackley and Xie, 2002; Xie and Tackley, 2004a).

7.12.5.2.1.3 Single-assemblage recipe

Another possible recipe is that OIB and MORB result from the same statistical distribution with their geochemical differences due entirely to melting processes, with the details of the melting processes varying in different proposals. In the original proposal (Allegre and Turcotte, 1986), enriched plums exist in a depleted matrix, and the higher heterogeneity of OIBs was ascribed to smaller sampling region and lower degrees of melt. More recent proposals tend to emphasize the heterogeneity of the entire mixture, with no~uniform residue. Based on modeling the evolution of the distribution of trace-element ratios in the differentiating continental crust–depleted mantle system, Kellogg (2004) proposed something of an inversion of this model, with depleted 'plums' consisting the residue from continental crust extraction, and a 'matrix' with a C/FOZO composition consisting of small-scale strips of everything else. Meibom and Anderson (2004) and Meibom et al. (2003) proposed that MORB and OIB derive from the same source region (statistical upper mantle assemblage (SUMA); the composition of the lower mantle was not specified) with the geochemical differences due to different statistical sampling of this highly heterogeneous source region – essentially, smaller melt fractions. A whole mantle consisting mainly of residue and strips of subducted crust was also favored by Helffrich and Wood (2001), who used various constraints to estimate plausible fractions of the different components that satisfy global heat budget constraints: for the mantle, their model contained between 10% and 24% recycled oceanic crust, with the rest mostly sterile mantle (i.e., zero heat-producing elements) and recycled continental material making up to 0.4%. The viability of this class of model is strongly supported by the modeling of Ito and Mahoney (2005a,b, 2006) who calculated the melting of a heterogeneous 4- or 5-component mantle and found that in most cases, OIBs and MORBs could be drawn from the same source distribution, although different OIBs require different proportions of the components.

The single-assemblage model has been the basis of various proposals discussed in the following sections.

7.12.5.2.2 Two-stage melting

It was proposed by Phipps Morgan and Morgan (1999), that the OIB and MORB signatures can be explained by two-stage melting of a heterogeneous 'plum pudding' source region. In this model, the OIB source is 'normal' mantle that includes lots of enriched basalt-pyroxenite veins, while the MORB source is produced from this by depletion associated with hot-spot melting. Thus, the initial stage of melting is associated with plumes producing OIBs with low-degree melting. The once-depleted plume material fills the asthenosphere, constituting the MORB source; subsequent second-stage melting underneath spreading centers produces the depleted MORB signature. A parameterized box model was presented to demonstrate the success of this model in explaining geochemical data. While this two-stage model is not widely accepted, it is useful to summarize the outcome of the parameterized model because the heterogeneous mixture of different components

could have wider applicability. The model mantle contains the following components: PRIM (primitive material), ORES (residue from OIB melting), MRES (residue from MORB melting), recycled components ZOIB, ZMORB, and ZCONT, and the continental crust. The rate of oceanic floor production is 2.7 km^2 $year^{-1}$ at the present day and increases in the past as (heat production)^2. By the end of the calculation, 95% of mantle has passed through MOR processing, and the proportions of the components are PRIM = 5%, ZCONT + ZOIB = 2%, ZMORB = 20%, ORES = 5%, and MRES = 62.5%, with continental crust being 0.5% . A Rb–Sr pseudoage of 1.3 Ga is obtained, which is much younger than whole-mantle pseudoisochron of 3.5 Ga because it represents the average time between OIB melting and MORB melting. Phipps Morgan and Morgan (1999) also presented an explanation for the Pb kappa conundrum (Th/U of MORB is 2.5, whereas lead ratios imply it should be 4) based on the different compatibilities of Th and U and for the heat–He imbalance based on He being outgassed mainly by hot spots and continental flood basalts (as previously suggested by Kellogg and Wasserburg (1990)), which leads to the prediction that atmospheric He concentrations should be elevated after flood basalt events.

7.12.5.2.3 Source statistics and component fractions

A series of papers (Anderson, 2000, 2001; Meibom and Anderson, 2004; Meibom et al., 2003; Meibom et al., 2002; Meibom et al., 2005) studied the statistical distributions of isotope ratios in MORB and OIB environments and found that they support the idea that the upper mantle consists of a multiple-length-scale mixture of different 'plums' with no uniform matrix, from long-term recycling of crustal and sedimentary components. Helium ratio distributions for the 'unfiltered' MORB dataset were plotted by Anderson (2000, 2001) and found to be drawn from the same statistical distribution as OIB helium ratios but with smaller standard deviation, which was argued to be related to averaging over larger sampling regions and degrees of melting consistent with the central limit theorem in which "independent averaging of samples from any given distribution will approach a Gaussian distribution with a variance that becomes smaller as the sample volume increases" (Meibom et al., 2002). This random sampling model was subsequently found to be consistent with the observed Gaussian distribution of $^{187}Os/^{188}Os$ in mantle-derived grains, implying random sampling between ancient radiogenic and unradiogenic mantle with relatively high degrees of partial melting (Meibom et al., 2002). The similarity between Os ratio distributions and He ratio distributions was discussed in Meibom et al. (2003) and argued to be evidence for the same processes acting on both systems, with Meibom and Anderson (2004) further analyzing ratios of Pb, Sr, and Nd, pointing out the Gaussian shape reflects the dominance of the homogenization during sampling process by partial melting and magma chamber mixing, rather than the nature of the source region. They argued that the dominant length scales are in the range 100 m to 100 km, rather than the cm to meter scale implicit in the original marble cake proposal (Allegre and Turcotte, 1986), and that therefore the efficiency of convective stirring is limited, with the dominant homogenization processes occurring during the melting and magmatic stages. There are also larger-scale variations related to different plate tectonic histories.

A length-scale constraint comes from the Os analysis of mantle-derived grains in Meibom et al. (2002): as a single grain is derived from about 1 m^3 of mantle, this is the minimum length scale of heterogeneity, and it is consistent with observations of Os variations in abyssal peridotites on length scales 10–100 m (Brandon et al., 2000; Parkinson et al., 1998). An age constraint comes from the least radiogenic Os ratio they found, which indicates a depletion age of 2.6 Gy.

Quantitative modeling of the distribution of isotope ratios by sampling a heterogeneous system was performed by Kellogg et al. (2002) using a new type of 'box model' that allows the distribution of isotope ratios to be calculated in addition to the mean values (Allegre and Lewin, 1995), as a function of melt fraction, sampling zone size, and evolution. Essentially, each melting event produces new 'subreservoirs,' which are stretched with time, evolve isotopically, and can be statistically sampled at a particular length scale. The results showed that indeed, as the sampling length scale or melt fraction is increased, isotope ratio distributions tend toward a Gaussian distribution, as discussed earlier. The method was used to successfully match Rb–Sr and Sm–Nd isotope ratio diagrams by differentiation and recycling in the upper mantle–continental crust system and subsequently extended to include the U–Th–Pb system in Kellogg (2004) and Kellogg et al. (2007). In the latter model, a lower continental crust was added, but the lower mantle remained primitive. It was found that matching lead isotopes required considering recycled MORB and again a change regarding relative U and Pb behavior at 2.3 Gyr before present. Preliminary results were obtained for a whole-mantle system including the oceanic differentiation cycle (Kellogg, 2004).

A simpler mathematical model of mantle isotopic variability was developed by Rudge et al. (2005). Their model considered the mantle to be a single reservoir that experiences ongoing melting events with a certain average time between remelting events. They found that the best fit to the isotopic scatter was obtained with remelting timescale of 1.4–2.4 Ga and melt fraction of around 0.5%, the latter of which is similar to Kellogg (2004) and Kellogg et al. (2002), although the latter study was intended to represent continental formation melting. This did not, however, fit Pb isotopic data.

7.12.5.2.4 Melting of a heterogeneous source

To understand the geochemical signature of erupted magmas and how it is related to the composition and heterogeneity of the source region, it is necessary to understand the trace- and major-element composition that is obtained when a heterogeneous mixture is partially melted.

When a source region that is heterogeneous in major-element chemistry is partially melted, the enriched pyroxenite material melts at a lower temperature than depleted residue, which means that low degrees of melting will preferentially sample the enriched material, while high degrees of melting will average over the whole assemblage (Sleep, 1984). Whether this can quantitatively account for the differences between MORB and OIB compositions has not, until recently, been quantitatively assessed. Note that this preferential sampling of enriched components is different from the proposal of Anderson (2001) and Meibom and Anderson (2004), in which the same material would be melted in MORB and OIB environments but with smaller melt fraction producing greater variability with the

same mean. Based on quantitative modeling of trace-element distributions, Kellogg et al. (2002) found that OIB and MORB cannot be produced from the same distribution simply by differences in sample volume (with smaller samples being more heterogeneous): differences in the melting processes and/or source regions are necessary.

If the depleted matrix (DM) and pyroxenite (PX) veins are at small enough length scales to be in thermal equilibrium, Phipps Morgan (2001) found that melting of the PX is greatly enhanced by heat flowing from the surrounding matrix. The details depend on the temperature–pressure gradient of PX solidus relative to DM solidus, consistent with other calculations (Hirschmann et al., 1999) and confirming the prediction of Sleep (1984).

Katz and Rudge (2011) considered the influence of the size and topology of fertile heterogeneities. They found an enhancement of melting by up to a factor of 2 for small ($<$1 km) heterogeneities, which is less than predicted by Sleep (1984) because of a more sophisticated petrologic and mathematical model.

Complicating interpretations of hot-spot melting is that different basalts from the same hot spot do not generate the same composition, but rather elongate tubelike fields (Hart et al., 1992), a traditional interpretation of which is a mixing line between two end-member components present in the source region. Phipps Morgan (1999) found, however, that such linear arrays are naturally generated by progressive melting of a heterogeneous source mixture and remain linear independent of the number of distinct isotopic components in the initial source mixture.

The earlier analyses are for equilibrium melting. Disequilibrium melting of a heterogeneous source was considered by Rudge et al. (2011), who found a generalization that reduces to the familiar batch and fractional melting models in the appropriate limits.

Some laboratory experiments emphasize the difference between melting MORB and melting pyroxenite. Kogiso et al. (2003) presented experiments showing that it is possible to generate OIB major-element composition by melting pyroxenite but not by melting MORB. Pertermann (2003) performed experiments of melting pyroxenite with a basaltic composition. They found that it starts melting 35–50 km deeper than peridotite and generates much more magma than the original amount of pyroxenite, estimating that there must be $\sim<2\%$ pyroxenite in the MORB source. One origin for pyroxenite is melting of MORB followed by reaction with surrounding peridotite (Sobolev et al., 2005).

Length-scale issues were considered by Kogiso et al. (2004). For the case of pyroxenite heterogeneities melting in a peridotite matrix, they investigated the conditions necessary for the melt to retain its isotopic heterogeneity. Segregation of the melt must occur faster than diffusive equilibration of the melt with surrounding solid. They calculated that pyroxenite bodies submeter to a few meters wide can retain and produce melts with distinctive geochemical signatures under typical mantle melting conditions.

7.12.5.2.5 Can OIB and MORB be produced by the same statistical assemblage?

From the considerations in the previous section, it is possible to consider the resulting signature when a heterogeneous mixture is melted in different environments – in particular, underneath spreading centers (MORB) versus underneath the lithosphere (OIBs). See **Chapter 7.10** for a review of some of the relevant observations and explanations. Different (OIB vs. MORB) melting environments will cause both major-element and trace-element variations, and correlation of major and trace elements in OIBs indicates that some of the isotopic variation can be explained by variations in melting (Jackson and Dasgupta, 2008).

7.12.5.2.5.1 Modeling basalt compositions

Calculations with up to five components were performed by Ito and Mahoney (2005a,b, 2006), who quantitatively modeled the different trace-element compositions that arise as a result of differing OIB and MORB environments. Two key differences arise between OIB and MORB melting environments as a result of the differing mantle flow. Whereas flow beneath a spreading center rises at a roughly constant velocity to the surface, the vertical flow caused by plumes impinging on the base of a rigid lithosphere decreases toward the base of the lithosphere tens of kilometers deep. Hence, the total amount of decompression melting that takes place in OIB environments is typically less, but even if the temperature is high enough to make it the same, the depth distribution of OIB melting is concentrated toward greater depths, whereas that of MORB melting is constant with depth. As the enriched components are assumed to melt at lower temperature (greater depth) due to their different lithologies and (sometimes) enrichment in volatiles, they make up a larger proportion of OIB-type melts, whereas MORB melting is diluted to a much larger extent by the depleted component, leading to a more enriched signature for OIBs. The details of the trace element are sensitive to the exact composition of enriched components in the melting zone, particularly for OIB melting.

Ito and Mahoney (2005a) studied melting of a heterogeneous mantle consisting of three components: DM = depleted mantle (similar to the DMM of Hart et al. (1992)), PX = pyroxenite (essentially subducted MORB and carrying the HIMU signature), and EM = enriched mantle (similar to EM1 and EM2). They calculated the trace-element signature (La/Sm, Sr, Nd, and Pb) of the resulting melt assuming perfect melt homogenization, as a function of the proportions of PX and EM (keeping DM fixed at 90%), the thickness of lithosphere, and the potential temperature. Their results reproduce many of the features observed in MORB and OIB data, particularly the difference between OIBs erupting over old (thick) seafloor and those erupting over young seafloor: over young seafloor, the compositions of hot-spot islands are less variable and overlap with MORB observations, whereas over old seafloor, trace-element compositions are more enriched and highly variable due to sensitivity to exact proportions of PX and EM, and temperature, in the source region. Their results are also able to reproduce some features of 'arrays' in ratio–ratio space, the trajectories of which are often curved and have corners due to the transition from melting one component to melting another. Because the observed isotope ratios of each hot spot typically cover a wide range (an elongated 'array'), the 'fit' must focus on fitting the scatter rather than precise values. Incomplete magma mixing during ascent, which is not treated, may cause melt batches to have more extreme isotopic signatures. Evidence for pressure being the most important sampling

parameter is in correlated major-element, trace-element, and isotope compositions in postglacial Icelandic basalts (Stracke et al., 2003), as well as other oceanic basalts (Jackson and Dasgupta, 2008).

This analysis was then applied to specific island chains in Ito and Mahoney (2005b), with the goal of answering the question of whether two distinct sources are needed to explain OIB and MORB. The authors used a simple grid search scheme to determine the range of mantle sources that can fit 13 specific hot-spot island chains based on correlations in Sr, Nd, Pb isotopes space and then calculated the isotopic composition you would get if melting each one below mid-ocean ridges. The variables are mantle temperature excess, mass fractions of components, and isotopic compositions of EM and PX. Most model OIB compositions produce MORB that is indistinguishable from observed MORB compositions, suggesting that the whole mantle could be made of the same statistical assemblage. The proportion of different components was different for different OIBs, indicating large-scale heterogeneity. However, some predicted MORB compositions lie outside the range of observed normal MORB data. The fit to MORB is optimal when DM fractions are 85–95%, with the rest being PX and EM. Regarding abundances of enriched components, on average, PX is slightly more abundant than EM (by a factor 1.17). Regarding the global budget issue, they calculated that if the upper bound for U and Th in continents and upper bound for U and Th in model are assumed, an additional primitive or otherwise enriched source is not necessary, though in that case, only 30% of the mantle + core heat budget is radiogenic.

This approach was extended to include He in Ito and Mahoney (2006), with the addition of a high $^{3}He/^{4}He$ component corresponding to C (Hanan and Graham, 1996) or FOZO (Hart et al., 1992; Hilton et al., 1999) and basically the same as HRDM (Stuart et al., 2003). It was found that OIB and MORB can arise out of the same heterogeneous source region if three conditions are met: (i) The high $^{3}He/^{4}He$ component starts melting deeper than DM, (ii) DM is $>=85\%$ of the mantle, and (iii) the concentration of He in C/FOZO is comparable to that in DM. The models are successful in matching the low variation in MORB $^{3}He/^{4}He$. The requirement for C/FOZO to have the same He concentration as DM implies that it is not primitive but has already melted and reinforces the findings of Class and Goldstein (2005) and earlier proposals and evidence discussed earlier (Albarède, 1998; Coltice and Ricard, 1999). In this view, C/FOZO keeps a high He ratio due to strong depletion in U and Th while retaining relatively high ^{3}He, perhaps in olivine-rich lithologies (Meibom et al., 2005; Parman et al., 2005), or by mixing with primitive material. Ito and Mahoney (2006) removed the requirement for this component to be distributed differently (e.g., located in the deep mantle) – it can be part of the general upper mantle assemblage as proposed by Meibom and Anderson (2004). Still, C/FOZO and DM are distinct components despite their similarity in major and trace elements other than He, so it needs to be established how such similar material can evolve two quite different He ratios.

Such a multicomponent melting treatment was incorporated into a 3-D numerical model of small-scale sublithospheric convection by Ballmer et al. (2010), who found that variations in the fractions of pyroxenite and enriched peridotitic components occurred naturally as a function of seafloor age and match observations of various chains in the western and southern Pacific. Additionally, complex age–distance relationships were displayed.

A simplified version of this, with only two components (10% enriched (EC) and 90% depleted (DC)), was applied to plume–lithosphere interaction underneath the Hawaiian chain (Bianco et al., 2008) and further developed to include the effects of preexisting small-scale convection by Ballmer et al. (2011), with some success at explaining isotopic heterogeneity as discussed in **Section 7.12.4.4**.

7.12.5.2.6 Basal Magma Ocean

A plausible mechanism that may have influenced the Earth's chemical and thermal evolution is the crystallization of a 'basal magma ocean' (Labrosse et al., 2007). In this scenario, the early magma ocean encompassing the entire mantle crystallized from the middle rather than the bottom as previously thought (Abe, 1997; Solomatov and Stevenson, 1993b), resulting in both shallow and BMOs. While the shallow magma ocean would have cooled and solidified rapidly as per previous analyses, the BMO would take a relatively long time to cool because heat transport is limited by the solid mantle. Thus, the deep mantle would have been partly molten for much of the Earth's history, with some pockets of remnant partial melt left at the present day in the form of ULVZs, as discussed in **Section 7.12.3.1.5**. There are several implications and consequences to such an evolution.

One is that heat-producing elements would partition into the melt; thus, the BMO would act as a reservoir of incompatible elements over the Earth's history, helping with the budget problems discussed in **Section 7.12.1.4**. This includes noble gases: Coltice et al. (2011) showed that under certain conditions, solid cumulates crystallizing from the BMO could explain the primitive-like He and Ne signatures in hot spots. A complexity here is that solid–melt partitioning depends on which solid minerals are present: if Ca-perovskite is present, then large-ion lithophile elements prefer to partition into that (Corgne et al., 2005; Hirose et al., 2004). However, Ca-perovskite is the first phase to melt.

Another is that as the BMO crystallized, the resulting solid would be relatively dense and might therefore remain above the CMB, being responsible for (or a component in) the dense piles discussed in other sections. Over time, fractional crystallization would have changed the composition of the remaining liquid and of the solid that progressively crystallized, such that the crystallizing solid would have become increasingly iron-rich and therefore dense.

The influence on the core would be profound early on. If the deep mantle is molten or mostly molten, then it will convect very rapidly and the TBL above the CMB will have only a very small temperature drop (Ulvrová et al., 2012), as does the TBL at the outside of the outer core. This means that the molten core and molten deep mantle will rapidly adjust to the same potential temperature and there will be a substantial TBL at the point where the mantle changes from mostly liquid to mostly solid, from which hot plumes could form. Furthermore, the large concentration of incompatible heat-producing elements in the BMO might lead to a slow cooling rate. The BMO would also strongly influence (essentially eliminate)

lateral variations of CMB heat flux. Another intriguing possibility is that a dynamo could be generated in the long-lived BMO rather than in the core itself (Ziegler and Stegman, 2013), something that might be able to drive a dynamo for about 2 billion years.

The main requirement for a BMO to be possible is that melt be denser than solid in the deepest mantle. A requirement for a complete magma ocean to solidify from the middle rather than from the bottom is that the melt adiabat intersects the liquidus in the midmantle, rather than at the base. Several researchers have been investigating whether these two requirements are true.

Regarding melt density, melt is normally less dense than solid if it has the same composition but could be denser than solid if the melt is enriched in Fe relative to the solid. Whether the latter is true for the lower mantle composition at CMB pressures is currently unclear. Nomura et al. (2011) found that for $(Mg,Fe)SiO_3$, melt is denser than the coexisting solid due to Fe partitioning, and de Koker et al. (2013) found this to be the case in the $MgO–SiO_2$ system, consistent with earlier experimental results for Mg_2SiO_4 (Mosenfelder et al., 2007). However, the experiments of Andrault et al. (2012) on a composition that includes aluminum indicate that Fe is insufficiently partitioned into the melt for it to be denser; the difference to the earlier results is attributed to the presence of aluminum.

Regarding melt adiabat and liquidus, it is now well established that for molten silicate, the Grüneisen parameter increases with pressure (de Koker et al., 2013; Mosenfelder et al., 2007, 2009; Stixrude et al., 2009), implying that the adiabatic gradient might increase with pressure, but for the liquidus (and solidus) curves, there is some disagreement. While several groups find a curved liquidus/solidus (de Koker et al., 2013; Fiquet et al., 2010; Mosenfelder et al., 2009; Stixrude et al., 2009), implying that a point of intersection might occur in midmantle, Andrault et al. (2011) found straight curves, which would intersect the melt adiabat at the CMB. These differences need to be resolved.

More research needs to be done to on these physical properties to understand the viability of a BMO.

7.12.5.2.7 Internal differentiation: 'Upside-down' and 'hybrid pyroxenite'

Melting-induced differentiation inside the mantle, that is, without the formation of crust, may also play an important role. One such type of internal differentiation was described by Sobolev et al. (2005): ascending MORB-derived eclogite partially melts; the melt starts migrating through host peridotite and reacts with it to form pyroxenite. Because pyroxenite has a higher solidus than MORB, it solidifies, although it could subsequently melt upon rising further. Such a mechanism is supported by measurements on Hawaiian rocks and was later used to calculate the amount of MORB in the source region for both MORBs and OIBs by Sobolev et al. (2007).

This mechanism was argued by Davies (2009) to be important in the Earth's trace-element budget and isotopic evolution. For this to be the case, some of the 'hybrid pyroxenite' must avoid remelting, thereby trapping trace elements in the residual zone, and be advected back into the deep mantle, where some of it should settle above the CMB due its higher density. This was subsequently applied to the noble gas budget (Davies, 2010), for which a key point is that during this internal differentiation, noble gases are not lost to the atmosphere, as they are with external differentiation (i.e., crustal formation).

The previously mentioned process could have happened throughout the Earth's history. Another internal differentiation process termed 'upside-down differentiation' is argued to have occurred in the early Earth by Lee et al. (2010). Due to the mantle being hotter than today, partial melting could have occurred much deeper in the upper mantle than it does today. Melt that formed deep in the hot, early upper mantle could have segregated downward rather than upward, then solidified, and sunk into the deepest mantle to form a primitive layer.

7.12.5.3 Water: Influence on Convection and the Transition Zone Water Filter

As discussed in **Section 7.12.1.10**, water is a potentially very important constituent in the Earth's mantle due to its effect on rheology (e.g., Hirth and Kohlstedt, 1996, although recently, this has been disputed Fei et al., 2013), and it is cycled between the interior and the biosphere/exosphere. In order for water to be carried into the deep mantle, it must first be incorporated into slabs and then escape dehydration processes as the slab sinks and warms; for a detailed review of these, see Faccenda (2014). Understanding dehydration of the slab requires detailed modeling incorporating complex phase diagrams (Arcay et al., 2005; Gerya and Yuen, 2003; Ruepke et al., 2004; van Keken et al., 2002, 2011). The latest such models (van Keken et al., 2011) predict that about 1/3 of the water bound in slabs makes it to 240 km depth, corresponding to about one ocean mass of water over the age of the Earth. Water that is lost from the slab plays an important role in transporting trace elements from the subducting slab to the mantle wedge region (Elliott, 2003) as well as causing melting that generates new crust. Recent analysis indicates that subducted seawater accounts for about 50% of the unbound water presently in the convecting mantle (Holland and Ballentine, 2006).

There have been several parameterized convection studies of the influence of water on global mantle evolution, all of which have found that it has a large influence. McGovern and Schubert (1989) found a negative feedback that stabilizes volatile content and cooling rate on a timescale of 100s Myr. Further developing this model, Sandu et al. (2011) also found a negative feedback that causes the final amount of water to be relatively independent of the initial amount and also the possibility that the overall balance switched from degassing to regassing as the Earth cooled. Crowley et al. (2011) analyzed the complex interaction between temperature and water feedbacks, finding possible states to be warming + degassing or cooling + regassing, with other combinations not compatible. In the future, these predictions of parameterized models should be verified in full convection models. So far, water cycling has only been incorporated into two global convection modeling studies (Fujita and Ogawa, 2013; Richard et al., 2002), neither of which attempted to deal with the complexities of subduction zones or to determine the influence of water on the Earth's evolution. The studies come to opposite conclusions regarding the distribution of water: Richard et al.

(2002) found it to be quite uniform (although they assumed more rapid hydrogen diffusion than might be realistic (Shaw et al., 2012)), while Fujita and Ogawa (2013) found it to be quite nonuniform.

It has also been proposed that water plays an important role in trace-element transport in the transition zone region. Due to different solubilities of water in different minerals, the transition zone between 410 and 660 km depth is capable of containing 10–30 times higher water concentrations than the mantle above or below it (Hirschmann, 2006a). This led to the idea (Bercovici and Karato, 2003) that some water may be trapped in the transition zone, and with it, incompatible trace elements. If so, this provides a way of maintaining a mantle that is layered in trace elements but still undergoing whole-mantle convection.

The high water content of enriched transition zone material rising through the 410 km discontinuity would cause it to partially melt, and the incompatible trace elements would enter the melt, leaving a depleted solid in the MORB source region. A requirement for this melt to be trapped is that it should be denser than the surrounding solid, in which case a melt layer would build up above 410 km, with the production of new melt being balanced by entrainment and solidification of melt near downwelling slabs. This mechanism applies to the slow passive upwelling return flow characteristic of a mainly internally heated system. Plumes from the water-poor lower mantle could, however, escape being stripped of their trace elements because they have a relatively low water content and move through the transition zone too quickly for water to diffuse into them. Thus, plumes in the upper mantle appear enriched relative to the filtered MORB. Material traveling in the opposite direction and entering the lower mantle across the 660 km discontinuity would likely be too cold to melt, but water could enter a free phase at grain boundaries and percolate back into the transition zone.

While appealing in its simplicity and ability to reconcile geochemical layering and whole-mantle convection, several aspects of the proposal require testing, as is presently happening. The existence of a partial melt layer above 410 km should be testable seismically, and some studies have indeed indicated such a zone (e.g., Courtier and Revenaugh, 2007; Jasbinsek and Dueker, 2007; Revenaugh and Sipkin, 1994; Song et al., 2004; Tauzin et al., 2010) or a possibly related abrupt change in anisotropy (Yuan and Beghein, 2013). An analysis of electrical conductivity of the relevant phases compared to that inferred in the mantle implies the predicted amount of water in the transition zone (Huang et al., 2005, 2006), although this has been questioned (Hirschmann, 2006b). Further discussions of water transport in the mantle can be found in Hirschmann (2006a).

Various research has been performed on dynamic aspects of this proposal, some of which is discussed in Karato et al. (2006). Leahy and Bercovici (2004) calculated the convective consequences of concentrating heat-producing elements in the transition zone and/or lower mantle, finding that the exact distribution has a minor influence on the flow except in the extreme case where all heat-producing elements are concentrated in the transition zone. Details of the spreading and entrainment (back into the transition zone) of the melt layer above 410 km were calculated by Leahy and Bercovici (2007) with the model extended to include reactive infiltration of hydrous melt into solid by Leahy and Bercovici (2010): melt layers with a thickness in excess of 10 km were found for a wide range of parameters. The influence of the melt–solid density crossover has a strong influence on the stability of such a melt layer (Youngs and Bercovici, 2009), with the crossover needing to be above the 410 km layer for it to be stable there. Water may be released from a slab that resides in the transition zone for some time (Richard et al., 2006), and the released water may cause small-scale convection above the slab due to buoyancy and low viscosity of the hydrated material (Richard and Bercovici, 2009).

Mineral properties such as the density of melt above 410 km depth, the partition coefficients or relevant trace elements at this pressure, and the diffusivity of hydrogen need to be better known. A geochemical prediction of the model is that at steady state, the average isotopic composition of the convecting upper mantle must be almost the same as that of the lower mantle, with modifications for radiogenic ingrowth, which implies that noble gas residence times in the upper mantle must be very long (C. J. Ballentine, personal communication, 2007).

7.12.5.4 Outlook and Future Directions

7.12.5.4.1 Can geochemical and geophysical observations be reconciled with current paradigms?

Recent advances in paradigms and modeling approaches bring us closer to an integrated model that is consistent with both geochemical and geophysical constraints. Indeed, the large number of uncertain parameters associated with the various physical processes that may be important (early differentiation possibly leading to layering (**Section 7.12.1.5.1**), crustal settling above the CMB (**Section 7.12.3.3.3**), partial layering around 660 km (**Section 7.12.3.3.4**), transition zone-water filter (**Section 7.12.5.3**), BMO (**Section 7.12.5.2.5**), internal differentiation (**Section 7.12.5.2.6**), reactions with the core (**Section 7.12.1.5.4**), shallow melting of a heterogeneous source (**Section 7.12.5.2.4**), and continental extraction and recycling (**Section 7.12.4.3**)) combined with the uncertain BSE composition (chondritic vs. nonchondritic (**Sections 7.12.1.2** and **7.12.1.6.3**) and noble gas abundances (**Section 7.12.5.1**)) likely leads to nonuniqueness and multiple solutions. Indeed, the discussions of the various geochemical paradoxes in earlier sections indicate multiple plausible hypotheses. More data and improved modeling will be key to resolving this.

Likely characteristics of a successful geodynamic–geochemical model are the following:

- Whole-mantle convection, perhaps with a dense layer near the base and some partial/intermittent stratification around 660 km depth.
- Almost all of the mantle (i.e., 95–99%) has been processed through mid-ocean ridge melting and so presumably been degassed of most primordial noble gases. This is consistent with the constraint of only ~50% degassing of radiogenic ^{40}Ar (**Section 7.12.1.6.1**).
- The mantle is compositionally heterogeneous at all scales, from thousands of km to centimeters (**Sections 7.12.1.7** and **7.12.2.7**).
- The turbulent (exponential) mixing regime, with a whole-mantle mixing time (to ~10 s cm scale lengths) of ~3 Gy at

present-day rates (Section 7.12.2.6.4). The upper mantle that is not much better mixed than the lower mantle (Sections 7.12.2.4.2 and 7.12.2.5). ('Upper mantle' mixing time estimates are meaningless in whole-mantle convection.)
- A residence time of ~6 Gy based on present-day rates (Sections 7.12.1.5.3 and 7.12.2.5), with faster processing in the past.

Less certain, but looking increasingly feasible are the following:

- Many OIBs may arise from a statistically similar heterogeneous source as MORB through differences in melting processes (Section 7.12.5.2.4). At the same time, there are some lateral and radial variations in the statistical assemblage.
- Some vertical stratification of incompatible trace elements due to either gravitational stratification of major elements (early enriched layer, subducted crust or other dense differentiation products settling above the CMB, and composition-dependent phase transitions around 660 km), effects of water (the transition zone water filter), or melting (BMO, internal differentiation processes).
- A nonchondritic BSE composition, resolving several geochemical paradoxes (Campbell and O'Neill, 2012).

Regarding possible dense material above the CMB, several of the previous mechanisms could have operated, leading to a heterogeneous assemblage of dense differentiated products – the basal mélange (BAM) (Tackley, 2012).

7.12.5.4.2 Quantitative modeling approaches

Recent years have seen the improvement of several modeling approaches that are capable of generating model geochemical data to compare to observations. These include improved 'box models' that treat the statistical distribution of isotope ratios as well as their mean values (Section 7.12.5.2.3), improved methods for calculating the trace-element signatures that arise from melting in different environments (Section 7.12.5.2.3), and numerical thermochemical convection models that track trace elements and make quantitative predictions about trace-element distributions and evolution as well as the seismic structure of the mantle (Sections 7.12.4.1–7.12.4.3). These different methods are complementary and can be combined to enhance understanding. Several modeling challenges remain in the future. The global numerical models presently do not include continent formation, because the relevant processes are poorly understood, and also cannot resolve the complex mass transport occurring in subduction zones, which is an important part of this. Global numerical models cannot treat the smaller scales, which must therefore be done using a statistical approach or local models.

Acknowledgments

The author thanks Chris Ballentine and Nicolas Coltice for thoughtful reviews on the original version of this chapter, David Bercovici for his editorial comments and organization, and Yanick Ricard for providing Figure 4.

References

Abe Y (1997) Thermal and chemical evolution of the terrestrial magma ocean. *Physics of the Earth and Planetary Interiors* 100: 27–39.
Agranier A, Blichert-Toft J, Graham DW, Debaille V, Schiano P, and Albarède F (2005) The spectra of isotopic heterogeneities along the Mid-Atlantic Ridge. *Earth and Planetary Science Letters* 238: 96–109.
Albarède F (1998) Time-dependent models of U–Th–He and K–Ar evolution and the layering of mantle convection. *Chemical Geology* 145: 413–429.
Albarède F (2001) Radiogenic ingrowth in systems with multiple reservoirs: Applications to the differentiation of the mantle-crust system. *Earth and Planetary Science Letters* 189: 59–73.
Albarède F (2005a) Helium feels the heat in Earth's mantle. *Science* 310: 1777–1778.
Albarède F (2005b) The survival of mantle geochemical heterogeneities. In: van der Hilst RD, Bass JD, Matas J, and Trampert J (eds.) *Earth's Deep Mantle: Structure, Composition, and Evolution*, pp. 27–46. Washington, DC: American Geophysical Union.
Albarede F (2008) Rogue mantle helium and neon. *Science* 319: 943–945.
Albarède F and van der Hilst RD (2002) Zoned mantle convection. *Philosophical Transactions of the Royal Society A* 360: 2569–2592.
Allegre CJ (1997) Limitation on the mass exchange between the upper and lower mantle: The evolving convection regime of the Earth. *Earth and Planetary Science Letters* 150: 1–6.
Allegre CJ, Brevart O, Dupre B, and Minster JF (1980) Isotopic and chemical effects produced in a continuously differentiating convecting Earth mantle. *Philosophical Transactions of the Royal Society A* 297: 447–477.
Allegre CJ, Hart SR, and Minster JF (1983a) Chemical structure and evolution of the mantle and continents determined by inversion of Nd and Sr isotopic data. I. Theoretical methods. *Earth and Planetary Science Letters* 66: 177–190.
Allegre CJ, Hofmann A, and O'Nions K (1996) The argon constraints on mantle structure. *Geophysical Research Letters* 23: 3555–3557.
Allegre CJ and Lewin E (1995) Isotopic systems and stirring times of the Earths mantle. *Earth and Planetary Science Letters* 136: 629–646.
Allegre CJ, Moreira M, and Staudacher T (1995) $^{4}He/^{3}He$ dispersion and mantle convection. *Geophysical Research Letters* 22: 2325–2328.
Allegre CJ, Sarda P, and Staudacher T (1993) Speculations about the cosmic origin of He and Ne in the interior of the earth. *Earth and Planetary Science Letters* 117: 229–233.
Allegre CJ, Staudacher T, Sarda P, and Kurz MD (1983b) Constraints on evolution of Earth's mantle from rare gas systematics. *Nature* 303: 762–766.
Allegre CJ and Turcotte DL (1986) Implications of a 2-component marble-cake mantle. *Nature* 323: 123–127.
Anderson DL (1993) He-3 from the mantle – Primordial signal or cosmic dust. *Science* 261: 170–176.
Anderson DL (1995) Lithosphere, asthenosphere, and perisphere. *Reviews of Geophysics* 33: 125–149.
Anderson DL (1998) A model to explain the various paradoxes associated with mantle noble gas geochemistry. *Proceedings of the National Academy of Sciences of the United States of America* 95: 9087–9092.
Anderson DL (2000) The statistics of helium isotopes along the global spreading ridge system and the central limit theorem. *Geophysical Research Letters* 27: 77–82.
Anderson DL (2001) A statistical test of the two reservoir model for helium isotopes. *Earth and Planetary Science Letters* 193: 77–82.
Anderson OL, Oda H, and Isaak D (1992) A model for the computation of thermal expansivity at high compression and high temperatures – MgO as an example. *Geophysical Research Letters* 19: 1987–1990.
Andrault D, Bolfan-Casanova N, Nigro GL, Bouhifd MA, Garbarino G, and Mezouar M (2011) Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history. *Earth and Planetary Science Letters* 304: 251–259.
Andrault D, Petitgirard S, Lo Nigro G, et al. (2012) Solid-liquid iron partitioning in Earth's deep mantle. *Nature* 487: 354–357.
Antolik M, Gu YJ, Ekstrom G, and Dziewonski AM (2003) J362D28: A new joint model of compressional and shear velocity in the Earth's mantle. *Geophysical Journal International* 153: 443–466.
Arcay D, Tric E, and Doin M-P (2005) Numerical simulations of subduction zones. Effect of slab dehydration on the mantle wedge dynamics. *Physics of the Earth and Planetary Interiors* 149: 133–153.

Armann M and Tackley PJ (2012) Simulating the thermo-chemical magmatic and tectonic evolution of Venus' mantle and lithosphere: Two-dimensional models. *Journal of Geophysical Research* 117: E12003, http://dx.doi.org/12012JE004231/12012JE004231.

Armstrong RL and Hein SM (1973) Computer simulation of Pb and Sr isotope evolution of the Earth's crust and upper mantle. *Geochimica et Cosmochimica Acta* 37: 1–18.

Asimow PD, Hirschmann MM, and Stolper EM (2001) Calculation of peridotite partial melting from thermodynamic models of minerals and melts, IV. Adiabatic decompression and the composition and mean properties of mid-ocean ridge basalts. *Journal of Petrology* 42: 963–998.

Ballentine CJ (2002) Geochemistry – Tiny tracers tell tall tales. *Science* 296: 1247–1248.

Ballentine CJ and Barford DN (2000) The origin of air-like noble gases in MORB and OIB. *Earth and Planetary Science Letters* 180: 39–48.

Ballentine CJ, Marty B, Lollar BS, and Cassidy M (2005) Neon isotopes constrain convection and volatile origin in the Earth's mantle. *Nature* 433: 33–38.

Ballentine CJ, van Keken PE, Porcelli D, and Hauri EH (2002) Numerical models, geochemistry and the zero-paradox noble-gas mantle. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 360: 2611–2631.

Ballmer MD, Ito G, van Hunen J, and Tackley PJ (2010) Small-scale sublithospheric convection reconciles geochemistry and geochronology of 'superplume' volcanism in the western and south Pacific. *Earth and Planetary Science Letters* 290: 224–232.

Ballmer MD, Ito G, van Hunen J, and Tackley PJ (2011) Spatial and temporal variability in Hawaiian hotspot volcanism induced by small-scale convection. *Nature Geoscience* 4: 457–460.

Barth MG, Rudnick RL, Horn I, et al. (2001) Geochemistry of xenolithic eclogites from west Africa, part I: A link between low MgO eclogites and Archean crust formation. *Geochimica et Cosmochimica Acta* 65: 1499–1527.

Batchelor GK (1952) The effect of homogeneous turbulence on material lines and surfaces. *Proceedings of the Royal Society A* 213: 349–366.

Batchelor GK (1959) Small-scale variation of convected quantities like temperature in a turbulent fluid. *Journal of Fluid Mechanics* 5: 113–133.

Becker TW, Kellogg JB, and O'Connell RJ (1999) Thermal constraints on the survival of primitive blobs in the lower mantle. *Earth and Planetary Science Letters* 171: 351–365.

Benz W and Cameron AGW (1990) Terrestrial effects of the giant impact. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 61–68. New York: Oxford University Press.

Bercovici D and Karato S (2003) Whole-mantle convection and the transition-zone water filter. *Nature* 425: 39–44.

Bianco TA, Ito G, van Hunen J, Ballmer MD, and Mahoney JJ (2008) Geochemical variation at the Hawaiian hot spot caused by upper mantle dynamics and melting of a heterogeneous plume. *Geochemistry, Geophysics, Geosystems* 9: Q11003.

Boehler R, Chopelas A, and Zerr A (1995) Temperature and chemistry of the core-mantle boundary. *Chemical Geology* 120: 199–205.

Bolton H and Masters G (2001) Travel times of P and S from the global digital seismic networks: Implications for the relative variation of P and S velocity in the mantle. *Journal of Geophysical Research: Solid Earth* 106: 13527–13540.

Bower DJ, Gurnis M, and Seton M (2013) Lower mantle structure from paleogeographically constrained dynamic Earth models. *Geochemistry, Geophysics, Geosystems* 14: 44–63.

Bower DJ, Wicks JK, Gurnis M, and Jackson JM (2011) A geodynamic and mineral physics model of a solid-state ultralow-velocity zone. *Earth and Planetary Science Letters* 303: 193–202.

Boyet M and Carlson RW (2005) 142Nd evidence for early (>4.53 Ga) global differentiation of the silicate Earth. *Science* 309: 576–581.

Boyet M and Carlson RW (2006) A new geochemical model for the Earth's mantle inferred from ^{146}Sm-^{142}Nd systematics. *Earth and Planetary Science Letters* 250: 254–268.

Brandenburg JP, Hauri EH, van Keken PE, and Ballentine CJ (2008) A multiple-system study of the geochemical evolution of the mantle with force-balanced plates and thermochemical effects. *Earth and Planetary Science Letters* 276: 1–13.

Brandenburg JP and van Keken PE (2007) Deep storage of oceanic crust in a vigorously convecting mantle. *Journal of Geophysical Research* 112: B06403. http://dx.doi.org/10.1029/2006JB004813.

Brandon A, Snow JE, Walker RJ, Morgan JW, and Mock TD (2000) ^{190}Pt-^{186}Os and ^{187}Re-^{187}Os systematics of abyssal peridotites. *Earth and Planetary Science Letters* 177: 319–335.

Brandon AD and Walker RJ (2005) The debate over core-mantle interaction. *Earth and Planetary Science Letters* 232: 211–225.

Brandon AD, Walker RJ, Morgan JW, Norman MD, and Prichard HM (1998) Coupled 186Os and 187Os evidence for core-mantle interaction. *Science* 280: 1570–1573.

Brandon AD, Walker RJ, Puchtel IS, Becker H, Humayun M, and Revillon S (2003) 186Os-187Os systematics of Gorgona Island komatiites: Implications for early growth of the inner core. *Earth and Planetary Science Letters* 206: 411–426.

Buffett BA, Garnero EJ, and Jeanloz R (2000) Sediments at the top of Earth's core. *Science* 290: 1338–1342.

Buffett BA, Huppert HE, Lister JR, and Woods AW (1992) Analytical model for solidification of the Earths core. *Nature* 356: 329–331.

Buffett BA, Huppert HE, Lister JR, and Woods AW (1996) On the thermal evolution of the Earth's core. *Journal of Geophysical Research* 101: 7989–8006.

Bull AL, McNamara AK, and Ritsema J (2009) Synthetic tomography of plume clusters and thermochemical piles. *Earth and Planetary Science Letters* 278: 152–162.

Bunge HP, Richards MA, and Baumgardner JR (1997) A sensitivity study of 3-dimensional spherical mantle convection at 10^8 Rayleigh number – Effects of depth-dependent viscosity, heating mode, and an endothermic phase change. *Journal of Geophysical Research* 102: 11991–12007.

Bunge HP, Richards MA, Lithgowbertelloni C, Baumgardner JR, Grand SP, and Romanowicz BA (1998) Time scales and heterogeneous structure in geodynamic Earth models. *Science* 280: 91–95.

Burke K, Steinberger B, Torsvik TH, and Smethurst MA (2008) Plume generation zones at the margins of large low shear velocity provinces on the core–mantle boundary. *Earth and Planetary Science Letters* 265: 49–60.

Burke K and Torsvik TH (2004) Derivation of Large Igneous Provinces of the past 200 million years from long-term heterogeneities in the deep mantle. *Earth and Planetary Science Letters* 227: 531–538.

Cabral RA, Jackson MG, Rose-Koga EF, et al. (2013) Anomalous sulphur isotopes in plume lavas reveal deep mantle storage of Archaean crust. *Nature* 496: 490–493.

Caffee MW, Hudson GB, Velsko C, Huss GR, Alexander EC, and Chivas AR (1999) Primordial noble gases from Earth's mantle: Identification of a primitive volatile component. *Science* 285: 2115–2118.

Campbell IH and O'Neill HS (2012) Evidence against a chondritic Earth. *Nature* 483: 553–558.

Caracas R and Cohen RE (2005) Effect of chemistry on the stability and elasticity of the perovskite and post-perovskite phases in the $MgSiO_3$–$FeSiO_3$–Al_2O_3 system and implications for the lowermost mantle. *Geophysical Research Letters* 32: L16310, http://dx.doi.org/16310.11029/12005gl023164.

Carlson RW (1987) Geochemical evolution of the crust and mantle. *Reviews of Geophysics* 25: 1011–1020.

Carlson RW (1994) Mechanisms of Earth differentiation – Consequences for the chemical-structure of the mantle. *Reviews of Geophysics* 32: 337–361.

Carlson RW and Boyet M (2008) Composition of the Earth's interior: The importance of early events. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 366: 4077–4103.

Caro G (2011) Early silicate earth differentiation. *Annual Review of Earth and Planetary Sciences* 39: 31–58.

Castillo P (1988) The Dupal anomaly as a trace of the upwelling lower mantle. *Nature* 336: 667–670.

Castle JC and van der Hilst RD (2003a) Searching for seismic scattering off mantle interfaces between 800 km and 2000 km depth. *Journal of Geophysical Research* 108: 2095. http://dx.doi.org/10.1029/2001JB000286.

Castle JC and van der Hilst RD (2003b) Using ScP precursors to search for mantle structures beneath 1800 km depth. *Geophysical Research Letters* 30: 1422. http://dx.doi.org/10.1029/2002GL016023.

Castro MC, Patriarche D, and Goblet P (2005) 2-D numerical simulations of groundwater flow, heat transfer and 4He transport – Implications for the He terrestrial budget and the mantle helium-heat imbalance. *Earth and Planetary Science Letters* 237: 893–910.

Chase CG (1981) Oceanic island Pb: Two-stage histories and mantle evolution. *Earth and Planetary Science Letters* 52: 277–284.

Chauvel CS, Goldstein SL, and Hofmann AW (1995) Hydration and dehydration of oceanic crust controls Pb evolution in the mantle. *Chemical Geology* 126: 65–75.

Chauvel C, Lewin E, Carpentier M, Arndt NT, and Marini J-C (2008) Role of recycled oceanic basalt and sediment in generating the Hf-Nd mantle array. *Nature Geoscience* 1: 64–67.

Chopelas A (1996) Thermal expansivity of lower mantle phases Mgo and $MgSiO_3$ perovskite at high-pressure derived from vibrational spectroscopy. *Physics of the Earth and Planetary Interiors* 98: 3–15.

Christensen U (1984) Instability of a hot boundary layer and initiation of thermo-chemical plumes. *Annales Geophysicae* 2: 311–319.

Christensen U (1988) Is subducted lithosphere trapped at the 670-km discontinuity? *Nature* 336: 462–463.

Christensen U (1989a) Mixing by time-dependent convection. *Earth and Planetary Science Letters* 95: 382–394.
Christensen UR (1989b) Models of mantle convection – One or several layers. *Philosophical Transactions of the Royal Society A* 328: 417–424.
Christensen U (1990) Mixing by time-dependent mantle convection – Reply. *Earth and Planetary Science Letters* 98: 408–410.
Christensen UR (1997) Influence of chemical buoyancy on the dynamics of slabs in the transition zone. *Journal of Geophysical Research* 102: 22435–22444.
Christensen UR and Hofmann AW (1994) Segregation of subducted oceanic crust in the convecting mantle. *Journal of Geophysical Research* 99: 19867–19884.
Christensen UR and Yuen DA (1984) The interaction of a subducting lithospheric slab with a chemical or phase-boundary. *Journal of Geophysical Research* 89: 4389–4402.
Christensen UR and Yuen DA (1985) Layered convection induced by phase transitions. *Journal of Geophysical Research* 90: 10291–10300.
Chudinovskikh L and Boehler R (2004) MgSiO/sub 3/ phase boundaries measured in the laser-heated diamond cell. *Earth & Planetary Science Letters* 219: 285–296.
Cizkova H, Cadek O, Van Den Berg AP, and Vlaar NJ (1999) Can lower mantle slab-like seismic anomalies be explained by thermal coupling between the upper and lower mantles? *Geophysical Research Letters* 26: 1501–1504.
Class C and Goldstein SL (2005) Evolution of helium isotopes in the Earth's mantle. *Nature* 436: 1107–1112.
Coltice N (2005) The role of convective mixing in degassing the Earth's mantle. *Earth and Planetary Science Letters* 234: 15–25.
Coltice N, Albarède F, and Gillet P (2000a) ^{40}K-^{40}Ar constraints on recycling continental crust into the mantle. *Science* 288: 845–847.
Coltice N, Ferrachat S, and Ricard Y (2000b) Box modeling the chemical evolution of geophysical systems: Case study of the Earth's mantle. *Geophysical Research Letters* 27: 1579–1582.
Coltice N, Marty B, and Yokochi R (2009) Xenon isotope constraints on the thermal evolution of the early Earth. *Chemical Geology* 266: 4–9.
Coltice N, Moreira M, Hernlund J, and Labrosse S (2011) Crystallization of a basal magma ocean recorded by Helium and Neon. *Earth and Planetary Science Letters* 308: 193–199.
Coltice N and Ricard Y (1999) Geochemical observations and one layer mantle convection. *Earth and Planetary Science Letters* 174: 125–137.
Coltice N and Ricard Y (2002) On the origin of noble gases in mantle plumes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 360: 2633–2648.
Coltice N and Schmalzl J (2006) Mixing times in the mantle of the early Earth derived from 2-D and 3-D numerical simulations of convection. *Geophysical Research Letters* 33: L23304. http://dx.doi.org/10.1029/2006GL027707.
Condie KC and Kröner A (2013) The building blocks of continental crust: Evidence for a major change in the tectonic setting of continental growth at the end of the Archean. *Gondwana Research* 23: 394–402.
Connolly JAD (2005) Computation of phase equilibria by linear programming: A tool for geodynamic modeling and an application to subduction zone decarbonation. *Earth and Planetary Science Letters* 236: 524–541.
Corgne A, Liebske C, Wood BJ, Rubie DC, and Frost DJ (2005) Silicate perovskite-melt partitioning of trace elements and geochemical signature of a deep perovskitic reservoir. *Geochimica et Cosmochimica Acta* 69: 485–496.
Corrsin S (1961) The reactant concentration spectrum in turbulent mixing with a first order reaction. *Journal of Fluid Mechanics* 11: 407–416.
Costin SO and Butler SL (2006) Modelling the effects of internal heating in the core and lowermost mantle on the Earth's magnetic history. *Physics of the Earth and Planetary Interiors* 157: 55–71.
Cottrell E and Kelley KA (2013) Redox heterogeneity in mid-ocean ridge basalts as a function of mantle source. *Science* 340: 1314–1317.
Courtier AM and Revenaugh J (2007) Deep upper-mantle melting beneath the Tasman and Coral Seas detected with multiple ScS reverberations. *Earth and Planetary Science Letters* 259: 66–76.
Creager KC and Jordan TH (1984) Slab penetration into the lower mantle. *Journal of Geophysical Research* 89: 3031–3049.
Creager KC and Jordan TH (1986) Slab penetration into the lower mantle beneath the Mariana and other island arcs of the Northwest Pacific. *Journal of Geophysical Research – B: Solid Earth and Planets* 91: 3573–3589.
Crisp JA (1984) Rates of magma emplacement and volcanic output. *Journal of Volcanology and Geothermal Research* 20: 177–211.
Crowley JW, Gérault M, and O'Connell RJ (2011) On the relative influence of heat and water transport on planetary dynamics. *Earth and Planetary Science Letters* 310: 380–388.
Cserepes L and Rabinowicz M (1985) Gravity and convection in a 2-layer mantle. *Earth and Planetary Science Letters* 76: 193–207.
Cserepes L, Rabinowicz M, and Rosembergborot C (1988) Three-dimensional infinite Prandtl number convection in one and two layers with implications for the Earth's gravity-field. *Journal of Geophysical Research – B: Solid Earth and Planets* 93: 12009–12025.
Dasgupta R (2013) Ingassing, storage, and outgassing of terrestrial carbon through geologic time. *Reviews in Mineralogy and Geochemistry* 75: 183–229.
Dasgupta R and Hirschmann MM (2010) The deep carbon cycle and melting in Earth's interior. *Earth and Planetary Science Letters* 298: 1–13.
Dasgupta R, Mallik A, Tsuno K, Withers AC, Hirth G, and Hirschmann MM (2013) Carbon-dioxide-rich silicate melt in the Earth's upper mantle. *Nature* 493: 211–215.
Davaille A (1999a) Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle. *Nature* 402: 756–760.
Davaille A (1999b) Two-layer thermal convection in miscible viscous fluids. *Journal of Fluid Mechanics* 379: 223–253.
Davaille A, Girard F, and Le BM (2002) How to anchor hotspots in a convecting mantle? *Earth and Planetary Science Letters* 203: 621–634.
Davies GF (1980) Thermal histories of convective Earth models and constraints on radiogenic heat production in the Earth. *Journal of Geophysical Research* 85: 2517–2530.
Davies GF (1981) Earth's neodymium budget and structure and evolution of the mantle. *Nature* 290: 208–213.
Davies GF (1983) Viscosity structure of a layered convecting mantle. *Nature* 301: 592–594.
Davies GF (1984) Geophysical and isotopic constraints on mantle convection – An interim synthesis. *Journal of Geophysical Research* 89: 6017–6040.
Davies GF (1988) Ocean bathymetry and mantle convection. 1. Large-scale flow and hotspots. *Journal of Geophysical Research – B: Solid Earth and Planets* 93: 10467–10480.
Davies GF (1990) Heat and mass transport in the early Earth. In: Newsome HE and Jones JH (eds.) *Origin of the Earth*, pp. 175–194. New York: Oxford University Press.
Davies GF (1992) On the emergence of plate-tectonics. *Geology* 20: 963–966.
Davies GF (1999) Geophysically constrained mantle mass flows and the ^{40}Ar budget: A degassed lower mantle? *Earth and Planetary Science Letters* 166: 149–162.
Davies GF (2002) Stirring geochemistry in mantle convection models with stiff plates and slabs. *Geochimica et Cosmochimica Acta* 66: 3125–3142.
Davies GF (2006) Depletion of the early Earth's upper mantle and the viability of early plate tectonics. *Earth and Planetary Science Letters* 243: 376–382.
Davies GF (2007) Controls on density stratification in the early mantle. *Geochemistry, Geophysics, Geosystems* 8: Q04006.
Davies GF (2008) Episodic layering of the early mantle by the 'basalt barrier' mechanism. *Earth and Planetary Science Letters* 275: 382–392.
Davies GF (2009) Reconciling the geophysical and geochemical mantles: Plume flows, heterogeneities, and disequilibrium. *Geochemistry, Geophysics, Geosystems* 10: Q10008.
Davies GF (2010) Noble gases in the dynamic mantle. *Geochemistry, Geophysics, Geosystems* 11: Q03005, http://dx.doi.org/02009GC002801/02009GC002801.
Davies DR, Goes S, Davies JH, Schuberth BSA, Bunge HP, and Ritsema J (2012) Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity. *Earth and Planetary Science Letters* 353: 253–269.
Davies GF and Gurnis M (1986) Interaction of mantle dregs with convection – Lateral heterogeneity at the core-mantle boundary. *Geophysical Research Letters* 13: 1517–1520.
de Koker N, Karki BB, and Stixrude L (2013) Thermodynamics of the MgO-SiO_2 liquid system in Earth's lowermost mantle from first principles. *Earth and Planetary Science Letters* 361: 58–63.
de Koker N, Steinle-Neumann G, and Vlcek V (2012) Electrical resistivity and thermal conductivity of liquid Fe alloys at high P and T, and heat flux in Earth's core. *Proceedings of the National Academy of Sciences of the United States of America* 109: 4070–4073.
Debaille V, O'Neill C, Brandon AD, et al. (2013) Stagnant-lid tectonics in early Earth revealed by 142Nd variations in late Archean rocks. *Earth and Planetary Science Letters* 373: 83–92.
Delano JW (2001) Redox history of the Earth's interior since similar to 3900 Ma: Implications for prebiotic molecules [Review]. *Origins of Life and Evolution of Biospheres* 31: 311–341.
Deparis V, Legros H, and Ricard Y (1995) Mass anomalies due to subducted slabs and simulations of plate motion since 200 My. *Physics of the Earth and Planetary Interiors* 89: 271–280.
Deschamps F, Kaminski E, and Tackley PJ (2011) A deep mantle origin for the primitive signature of ocean island basalt. *Nature Geoscience* 4: 879–882.

Deschamps F and Tackley PJ (2008) Searching for models of thermo-chemical convection that explain probabilistic tomography I. Principles and influence of rheological parameters. *Physics of the Earth and Planetary Interiors* 171: 357–373.

Deschamps F and Tackley PJ (2009) Searching for models of thermo-chemical convection that explain probabilistic tomography II – Influence of physical and compositional parameters. *Physics of the Earth and Planetary Interiors* 176: 1–18.

Deschamps F, Trampert J, and Tackley PJ (2007) Thermo-chemical structure of the lower mantle: Seismological evidence and consequences for geodynamics. In: Yuen DA, Maruyama S, Karato SI, and Windley BF (eds.) *Superplume: Beyond Plate Tectonics*, pp. 293–320. Dordrecht: Springer.

Dhuime B, Hawkesworth CJ, Cawood PA, and Storey CD (2012) A change in the geodynamics of continental growth 3 billion years ago. *Science* 335: 1334–1336.

Dixon ET, Honda M, McDougall I, Campbell IH, and Sigurdsson I (2000) Preservation of near-solar isotopic ratios in Icelandic basalts. *Earth and Planetary Science Letters* 180: 309–324.

Dobson DP and Brodholt JP (2005) Subducted banded iron formations as a source of ultralow-velocity zones at the core-mantle boundary. *Nature* 434: 371–374.

Drake MJ and Righter K (2002) Determining the composition of the Earth [Review]. *Nature* 416: 39–44.

du Vignaux NM and Fleitout L (2001) Stretching and mixing of viscous blobs in Earth's mantle. *Journal of Geophysical Research: Solid Earth* 106: 30893–30908.

Dubrovinsky L, Annersten H, Dubrovinskaia N, et al. (2001) Chemical interaction of Fe and Al_2O_3 as a source of heterogeneity at the Earth's core-mantle boundary. *Nature* 412: 527–529.

Dubrovinsky L, Dubrovinskaia N, Langenhorst F, et al. (2003) Iron-silica interaction at extreme conditions and the electrically conducting layer at the base of Earth's mantle. *Nature* 421: 58–61.

Dupeyrat L, Sotin C, and Parmentier EM (1995) Thermal and chemical convection in planetary mantles. *Journal of Geophysical Research – B: Solid Earth and Planets* 100: 497–520.

Dupre B and Allegre CJ (1983) Pb-Sr variation in Indian Ocean basalts and mixing phenomena. *Nature* 303: 142–146.

Dziewonski AM, Lekic V, and Romanowicz BA (2010) Mantle anchor structure: An argument for bottom up tectonics. *Earth and Planetary Science Letters* 299: 69–79.

Elkins-Tanton LT (2008) Linked magma ocean solidification and atmospheric growth for Earth and Mars. *Earth and Planetary Science Letters* 271: 181–191.

Elkins-Tanton LT, Parmentier EM, and Hess PC (2003) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteoritics and Planetary Science* 38: 1753–1771.

Elliott TR (2003) Tracers of the Slab. In: Eiler JM (ed.) *Inside the Subduction Factory*, pp. 23–45. Washington, DC: American Geophysical Union.

Elliott T, Zindler A, and Bourdon B (1999) Exploring the kappa conundrum: The role of recycling in the lead isotope evolution of the mantle. *Earth and Planetary Science Letters* 169: 129–145.

Faccenda M (2014) Water in the slab: A trilogy. *Tectonophysics* 614: 1–30.

Farley KA, Natland JH, and Craig H (1992) Binary mixing of enriched and undegassed (primitive?) mantle components (He, Sr, Nd, Pb) in samoan lavas. *Earth and Planetary Science Letters* 111: 183–199.

Farley KA and Neroda E (1998) Noble gases in the Earth's mantle. *Annual Review of Earth and Planetary Sciences* 26: 189–218.

Farnetani CG (1997) Excess temperature of mantle plumes – The role of chemical stratification across D″. *Geophysical Research Letters* 24: 1583–1586.

Farnetani CG and Hofmann AW (2009) Dynamics and internal structure of a lower mantle plume conduit. *Earth and Planetary Science Letters* 282: 314–322.

Farnetani CG and Hofmann AW (2010) Dynamics and internal structure of the Hawaiian plume. *Earth and Planetary Science Letters* 295: 231–240.

Farnetani CG, Hofmann AW, and Class C (2012) How double volcanic chains sample geochemical anomalies from the lowermost mantle. *Earth and Planetary Science Letters* 359–360: 240–247.

Farnetani CG, Legras B, and Tackley PJ (2002) Mixing and deformations in mantle plumes. *Earth and Planetary Science Letters* 196: 1–15.

Farnetani DG and Richards MA (1995) Thermal entrainment and melting in mantle plumes. *Earth and Planetary Science Letters* 136: 251–267.

Farnetani CG and Samuel H (2003) Lagrangian structures and stirring in the Earth's mantle. *Earth and Planetary Science Letters* 206: 335–348.

Farnetani CG and Samuel H (2005) Beyond the thermal plume paradigm. *Geophysical Research Letters* 32: L07311. http://dx.doi.org/10.1029/2005GL022360.

Fei H, Wiedenbeck M, Yamazaki D, and Katsura T (2013) Small effect of water on upper-mantle rheology based on silicon self-diffusion coefficients. *Nature* 498: 213–215.

Ferrachat S and Ricard Y (1998) Regular vs. chaotic mantle mixing. *Earth and Planetary Science Letters* 155: 75–86.

Ferrachat S and Ricard Y (2001) Mixing properties in the Earth's mantle: Effects of the viscosity stratification and of oceanic crust segregation. *Geochemistry, Geophysics, Geosystems* 2: 1013. http://dx.doi.org/10.1029/2000GC000092.

Fiquet G, Auzende AL, Siebert J, et al. (2010) Melting of peridotite to 140 gigapascals. *Science* 329: 1516–1518.

Fleitout L, Mambole A, and Christensen U (2000) Phase changes around 670 km depth and segregation in the Earth's mantle. *Eos, Transactions American Geophysical Union* 81, Fall Meeting Suppl., Abstract T12E-11.

Foley SF, Buhre S, and Jacob DE (2003) Evolution of the Archaean crust by delamination and shallow subduction. *Nature* 421: 249–252.

Forte AM and Mitrovica JX (2001) Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data. *Nature* 410: 1049–1056.

Fujita K and Ogawa M (2009) Basaltic accumulation instability and chaotic plate motion in the earliest mantle inferred from numerical experiments. *Journal of Geophysical Research* 114: B10402.

Fujita K and Ogawa M (2013) A preliminary numerical study on water-circulation in convecting mantle with magmatism and tectonic plates. *Physics of the Earth and Planetary Interiors* 216: 1–11.

Fukao Y (1992) Seismic tomogram of the earths mantle – Geodynamic implications. *Science* 258: 625–630.

Fukao Y, Obayashi M, Inoue H, and Nenbai M (1992) Subducting slabs stagnant in the mantle transition zone. *Journal of Geophysical Research: Solid Earth* 97: 4809–4822.

Gable CW, O'Connell RJ, and Travis BJ (1991) Convection in 3 dimensions with surface plates – Generation of toroidal flow. *Journal of Geophysical Research* 96: 8391–8405.

Gaherty JB and Hager BH (1994) Compositional vs thermal buoyancy and the evolution of subducted lithosphere. *Geophysical Research Letters* 21: 141–144.

Galer SJG and O'Nions RK (1986) Magmagenesis and the mapping of chemical and isotopic variations in the mantle. *Chemical Geology* 56: 45–61.

Garnero EJ (2004) A new paradigm for Earth's core-mantle boundary. *Science* 304: 834–836. http://dx.doi.org/10.1126/science.1097849.

Garnero EJ and Helmberger DV (1995) A very slow basal layer underlying large-scale low-velocity anomalies in the lower mantle beneath the pacific – Evidence from core phases. *Physics of the Earth and Planetary Interiors* 91: 161–176.

Garnero EJ and Helmberger DV (1996) Seismic detection of a thin laterally varying boundary-layer at the base of the mantle beneath the central Pacific. *Geophysical Research Letters* 23: 977–980.

Gerya TV and Yuen DA (2003) Rayleigh-Taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones. *Earth and Planetary Science Letters* 212: 47–62.

Glatzmaier GA and Schubert G (1993) 3-Dimensional spherical-models of layered and whole mantle convection. *Journal of Geophysical Research* 98: 21969–21976.

Goarant F, Guyot F, Peyronneau J, and Poirier JP (1992) High-pressure and high-temperature reactions between silicates and liquid iron alloys, in the diamond anvil cell, studied by analytical electron microscopy. *Journal of Geophysical Research* 97: 4477–4487.

Gomi H, Ohta K, Hirose K, et al. (2013) The high conductivity of iron and thermal evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 224: 88–103.

Gonnermann HM, Manga M, and Jellinek AM (2002) Dynamics and longevity of an initially stratified mantle. *Geophysical Research Letters* 29: 1399, http://dx.doi.org/1310.1029/2002GL014851.

Gonnermann HM and Mukhopadhyay S (2007) Non-equilibrium degassing and a primordial source for helium in ocean-island volcanism. *Nature* 449: 1037–1040.

Gonnermann HM and Mukhopadhyay S (2009) Preserving noble gases in a convecting mantle. *Nature* 459: 560–563.

Graham DW (2002) Noble gas isotope geochemistry of mid-ocean ridge and ocean island basalts; characterization of mantle source reservoirs. In: Porcelli D, Wieler R, and Ballentine CJ (eds.) *Noble Gases in Geochemistry and Cosmochemistry, Reviews in Mineralogy and Geochemistry*, pp. 247–318. Washington, DC: Mineralogical Society of America.

Graham DW, Blichert-Toft J, Russo CJ, Rubin KH, and Albarède F (2006) Cryptic striations in the upper mantle revealed by hafnium isotopes in southeast Indian ridge basalts. *Nature* 440: 199–202.

Graham D, Lupton J, Albarède F, and Condomines M (1990) Extreme temporal homogeneity of Helium isotopes at Piton de la Fournaise, Reunion Island. *Nature* 347: 545–548.

Graham DW, Lupton JE, Spera FJ, and Christie DM (2001) Upper-mantle dynamics revealed by helium isotope variations along the Southeast Indian Ridge. *Nature* 409: 701–703.

Grand SP (1994) Mantle shear structure beneath the america and surrounding oceans. *Journal of Geophysical Research: Solid Earth* 99: 11591–11621.
Griffiths RW (1986) Dynamics of mantle thermals with constant buoyancy or anomalous internal heating. *Earth and Planetary Science Letters* 78: 435–446.
Guignot N and Andrault D (2004) Equations of state of Na-K-Al host phases and implications for MORB density in the lower mantle. *Physics of the Earth and Planetary Interiors* 143–144: 107–128.
Gurnis M (1986a) The effects of chemical density differences on convective mixing in the Earth's mantle. *Journal of Geophysical Research* 91: 1407–1419.
Gurnis M (1986b) Quantitative bounds on the size spectrum of isotopic heterogeneity within the mantle. *Nature* 323: 317–320.
Gurnis M (1986c) Stirring and mixing in the mantle by plate-scale flow – Large persistent blobs and long tendrils coexist. *Geophysical Research Letters* 13: 1474–1477.
Gurnis M and Davies GF (1986a) The effect of depth-dependent viscosity on convective mixing in the mantle and the possible survival of primitive mantle. *Geophysical Research Letters* 13: 541–544.
Gurnis M and Davies GF (1986b) Mixing in numerical-models of mantle convection incorporating plate kinematics. *Journal of Geophysical Research – B: Solid Earth and Planets* 91: 6375–6395.
Hager BH, Clayton RW, Richards MA, Comer RP, and Dziewonski AM (1985) Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* 313: 541–546.
Hager BH and Richards MA (1989) Long-wavelength variations in Earths geoid – Physical models and dynamical implications. *Philosophical Transactions of the Royal Society A* 328: 309–327.
Hanan BB and Graham DW (1996) Lead and helium isotope evidence from oceanic basalts for a common deep source of mantle plumes. *Science* 272: 991–995.
Hansen U and Yuen DA (1988) Numerical simulations of thermal-chemical instabilities at the core mantle boundary. *Nature* 334: 237–240.
Hansen U and Yuen DA (1989) Dynamical influences from thermal-chemical instabilities at the core-mantle boundary. *Geophysical Research Letters* 16: 629–632.
Hansen U and Yuen DA (2000) Extended-Boussinesq thermal-chemical convection with moving heat sources and variable viscosity. *Earth and Planetary Science Letters* 176: 401–411.
Harrison TM (2009) The hadean crust: Evidence from >4 Ga zircons. *Annual Review of Earth and Planetary Sciences* 37: 479–505.
Harrison TM, Blichert-Toft J, Mueller W, Albarade F, Holden P, and Mojzsis SJ (2005) Heterogeneous hadean hafnium: Evidence of continental crust at 4.4 to 4.5 Gyr. *Science* 310: 1947–1950.
Harrison D, Burnard P, and Turner G (1999) Noble gas behaviour and composition in the mantle: Constraints from the Iceland plume. *Earth and Planetary Science Letters* 171: 199–207.
Hart SR (1984) A large-scale isotope anomaly in the Southern Hemisphere mantle. *Nature* 309: 753–757.
Hart SR, Hauri EH, Oschmann JA, and Whitehead JA (1992) Mantle plumes and entrainment: Isotopic evidence. *Science* 256: 517–520.
Hauri E (2002) SIMS analysis of volatiles in silicate glasses: 2. Isotopes and abundances in Hawaiian melt inclusions. *Chemical Geology* 183: 115–141.
Hauri EH, Whitehead JA, and Hart SR (1994) Fluid dynamic and geochemical aspects of entrainment in mantle plumes. *Journal of Geophysical Research, Solid Earth* 99: 24275–24300.
Hawkesworth CJ and Kemp AIS (2006) Evolution of the continental crust. *Nature* 443: 811–817.
He Y and Wen L (2009) Structural features and shear-velocity structure of the "Pacific Anomaly". *Journal of Geophysical Research* 114: B02309.
Heber VS, Brooker RA, Kelley SP, and Wood BJ (2007) Crystal-melt partitioning of noble gases (helium, neon, argon, krypton, and xenon) for olivine and clinopyroxene. *Geochimica et Cosmochimica Acta* 71: 1041–1061.
Hedlin MAH and Shearer PM (2000) An analysis of large-scale variations in small-scale mantle heterogeneity using global seismographic network recordings of precursors to PKP. *Journal of Geophysical Research* 105: 13655–13673.
Hedlin MAH, Shearer PM, and Earle PS (1997) Seismic evidence for small-scale heterogeneity throughout the Earth's mantle. *Nature* 387: 145–150.
Helffrich GR and Wood BJ (2001) The Earth's mantle [Review]. *Nature* 412: 501–507.
Hernlund JW and Jellinek AM (2010) Dynamics and structure of a stirred partially molten ultralow-velocity zone. *Earth and Planetary Science Letters* 296: 1–8.
Hernlund JW and Tackley PJ (2007) Some dynamical consequences of partial melting in Earth's deep mantle. *Physics of the Earth and Planetary Interiors* 162: 149–163.
Hernlund JW and Tackley P (2008) Modeling mantle convection in the spherical annulus. *Physics of the Earth and Planetary Interiors* 171: 48–54.
Hernlund JW, Thomas C, and Tackley PJ (2005) A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle. *Nature* 434: 882–886.
Herzberg C and Rudnick R (2012) Formation of cratonic lithosphere: An integrated thermal and petrological model. *Lithos* 149: 4–15.
Hilton DR, Gronvold K, Macpherson CG, and Castillo PR (1999) Extreme $^{3}He/^{4}He$ ratios in northwest Iceland: Constraining the common component in mantle plumes. *Earth and Planetary Science Letters* 173: 53–60.
Hilton DR and Porcelli D (2003) Noble gases as mantle tracers. In: Carlson RW (ed.) *Treatise on Geochemistry. The Mantle and Core*, vol. 2, pp. 277–318. Amsterdam: Elsevier.
Hirose K (2002) Phase transitions in pyrolitic mantle around 670-km depth: Implications for upwelling of plumes from the lower mantle. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000597.
Hirose K and Kushiro I (1993) Partial melting of dry peridotites at high pressures: Determination of compositions of melts segregated from peridotite using aggregates of diamond. *Earth and Planetary Science Letters* 114: 477–489.
Hirose K, Shimizu N, van Westeren W, and Fei Y (2004) Trace element partitioning in Earth's lower mantle and implications for geochemical consequences of partial melting at the core-mantle boundary. *Physics of the Earth and Planetary Interiors* 146: 249–260.
Hirose K, Takafuji N, Sata N, and Ohishi Y (2005) Phase transition and density of subducted MORB crust in the lower mantle. *Earth and Planetary Science Letters* 237: 239–251.
Hirschmann MM (2006a) Water, melting, and the deep Earth H_2O cycle. *Annual Review of Earth and Planetary Sciences* 34: 629–653.
Hirschmann MM (2006b) A wet mantle conductor? *Nature* 439: E3.
Hirschmann MM, Asimow PD, Ghiorso MS, and Stolper EM (1999) Calculation of peridotite partial melting from thermodynamic models of minerals and melts. III. Controls on isobaric melt production and the effect of water on melt production. *Journal of Petrology* 40: 831–851.
Hirth G and Kohlstedt DL (1996) Water in the oceanic upper-mantle – Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth and Planetary Science Letters* 144: 93–108.
Hoffman NRA and McKenzie DP (1985) The destruction of geochemical heterogeneities by differential fluid motions during mantle convection. *Geophysical Journal of the Royal Astronomical Society* 82: 163–206.
Hofmann AW (1988) Chemical differentiation of the Earth: The relationship between mantle, continental crust, and oceanic crust. *Earth and Planetary Science Letters* 90: 297–314.
Hofmann AW (1997) Mantle geochemistry: The message from oceanic volcanism. *Nature* 385: 219–229.
Hofmann AW (2003) Sampling mantle heterogeneity through oceanic basalts: Isotopes and trace elements. In: Carlson RW (ed.) *Treatise on Geochemistry*, pp. 61–101. Oxford: Elsevier.
Hofmann AW, Farnetani CG, Spiegelman M, and Class C (2011) Displaced helium and carbon in the Hawaiian plume. *Earth and Planetary Science Letters* 312: 226–236.
Hofmann AW and Hart SR (1978) An assessment of local and regional isotopic equilibrium in the mantle. *Earth and Planetary Science Letters* 38: 44–62.
Hofmann AW and White WM (1982) Mantle plumes from ancient oceanic-crust. *Earth and Planetary Science Letters* 57: 421–436.
Hoink T, Schmalzl J, and Hansen U (2006) Dynamics of metal-silicate separation in a terrestrial magma ocean. *Geochemistry, Geophysics, Geosystems* 7: Q09008. http://dx.doi.org/10.1029/2006GC001268.
Holland HD (1984) *The Chemical Evolution of the Atmosphere and Oceans*. Princeton, NJ: Princeton University Press.
Holland HD (1994) Early proterozoic atmospheric change. In: Bengtson S (ed.) *Early Life on Earth*, pp. 237–244. New York: Columbia University Press.
Holland G and Ballentine CJ (2006) Seawater subduction controls the heavy noble gas composition of the mantle. *Nature* 441: 186–191.
Honda M, McDougall I, Patterson DB, Doulgeris A, and Clague DA (1999) Noble gases in submarine pillow basalt glasses from Loihi and Kilauea, Hawaii: A solar component in the Earth. *Geochimica et Cosmochimica Acta* 57: 859–874.
Hopp J and Trieloff M (2005) Refining the noble gas record of the Reunion mantle plume source: Implications on mantle geochemistry. *Earth and Planetary Science Letters* 240: 573–588.
Huang J and Davies GF (2007a) Geochemical processing in a three-dimensional regional spherical shell model of mantle convection. *Geochemistry, Geophysics, Geosystems* 8: Q11006. http://dx.doi.org/10.1029/2007GC001625.
Huang J and Davies GF (2007b) Stirring in three-dimensional mantle convection models and implications for geochemistry: 2. Heavy tracers. *Geochemistry, Geophysics, Geosystems* 8: Q07004.

Huang J and Davies GF (2007c) Stirring in three-dimensional mantle convection models and implications for geochemistry: Passive tracers. *Geochemistry, Geophysics, Geosystems* 8: Q03017.

Huang X, Xu Y, and Karato S (2005) Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite. *Nature* 434: 746–749.

Huang X, Xu Y, and Karato S (2006) A wet mantle conductor? Reply. *Nature* 439: E3–E4.

Humayun M, Qin L, and Norman MD (2004) Geochemical evidence for excess iron in the mantle beneath Hawaii. *Science* 306: 91–94.

Hunt DL and Kellogg LH (2001) Quantifying mixing and age variations of heterogeneities in models of mantle convection: Role of depth-dependent viscosity. *Journal of Geophysical Research* 106: 6747–6760.

Irifune T and Ringwood AE (1993) Phase-transformations in subducted oceanic-crust and buoyancy relationships at depths of 600-800 km in the mantle. *Earth and Planetary Science Letters* 117: 101–110.

Ishii M and Tromp J (1999) Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle. *Science* 285: 1231–1235.

Ishii M and Tromp J (2004) Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes. *Physics of the Earth and Planetary Interiors* 146: 113–124.

Ito G and Mahoney JJ (2005a) Flow and melting of a heterogeneous mantle: 1. Method and importance to the geochemistry of ocean island and mid-ocean ridge basalts. *Earth and Planetary Science Letters* 230: 29–46.

Ito G and Mahoney JJ (2005b) Flow and melting of a heterogeneous mantle: 2. Implications for a chemically nonlayered mantle. *Earth and Planetary Science Letters* 230: 47–63.

Ito G and Mahoney JJ (2006) Melting a high $^{3}He/^{4}He$ source in a heterogeneous mantle. *Geochemistry, Geophysics, Geosystems* 7: Q05010, http://dx.doi.org/05010.01029/02005GC001158.

Iwamori H and Albarede F (2008) Decoupled isotopic record of ridge and subduction zone processes in oceanic basalts by independent component analysis. *Geochemistry, Geophysics, Geosystems* 9: Q04033.

Iwamori H, Albaréde F, and Nakamura H (2010) Global structure of mantle isotopic heterogeneity and its implications for mantle differentiation and convection. *Earth and Planetary Science Letters* 299: 339–351.

Jackson MG, Carlson RW, Kurz MD, Kempton PD, Francis D, and Blusztajn J (2010) Evidence for the survival of the oldest terrestrial mantle reservoir. *Nature* 466: 853–856.

Jackson MG and Dasgupta R (2008) Compositions of HIMU, EM1, and EM2 from global trends between radiogenic isotopes and major elements in ocean island basalts. *Earth and Planetary Science Letters* 276: 175–186.

Jacobsen SB and Wasserburg GJ (1979) The mean age of mantle and crustal reservoirs. *Journal of Geophysical Research* 84: 7411–7427.

Jasbinsek J and Dueker K (2007) Ubiquitous low-velocity layer atop the 410-km discontinuity in the northern Rocky Mountains. *Geochemistry, Geophysics, Geosystems* 8: Q10004.

Jellinek AM and Manga M (2002) The influence of a chemical boundary layer on the fixity and lifetime of mantle plumes. *Nature* 418: 760–763.

Jochum K, Hofmann AW, Ito E, Seufert HM, and White WM (1983) K, U and Th in mid-ocean ridge basalt glasses and heat production, K/U and K/Rb in the mantle. *Nature* 306: 431–436.

Johnson TE, Brown M, Kaus BJP, and VanTongeren JA (2014) Delamination and recycling of Archaean crust caused by gravitational instabilities. *Nature Geoscience* 7: 47–52.

Jordan TH, Puster P, Glatzmaier GA, and Tackley PJ (1993) Comparisons between seismic Earth structures and mantle flow models based on radial correlation functions. *Science* 261: 1427–1431.

Jull M and Kelemen PB (2001) On the conditions for lower crustal convective instability. *Journal of Geophysical Research* 106: 6423.

Kameyama M, Fujimoto H, and Ogawa M (1996) A thermo-chemical regime in the upper mantle in the early Earth inferred from a numerical model of magma-migration in a convecting upper mantle. *Physics of the Earth and Planetary Interiors* 94: 187–215.

Kanda RVS and Stevenson DJ (2006) Suction mechanism for iron entrainment into the lower mantle. *Geophysical Research Letters* 33: L02310. http://dx.doi.org/10.1029/2005GL025009.

Kaneshima S and Helffrich G (1999) Dipping low-velocity layer in the mid-lower mantle: Evidence for geochemical heterogeneity. *Science* 283: 1888–1891.

Kaneshima S and Helffrich G (2009) Lower mantle scattering profiles and fabric below Pacific subduction zones. *Earth and Planetary Science Letters* 282: 234–239.

Karato S (1995) Interaction of chemically stratified subducted oceanic lithosphere with the 660 km discontinuity. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences* 71: 203–207.

Karato S, Bercovici D, Leahy G, Richard G, and Jing Z (2006) The transition zone water filter model for global material circulation: Where do we stand? In: Jacobsen SD and van der Lee S (eds.) *Earth's Deep Water Cycle*. *AGU Monograph Series*, vol. 168, pp. 289–314. Washington, DC: American Geophysical Union.

Karato S and Murthy VR (1997) Core formation and chemical-equilibrium in the Earth. 1. Physical considerations. *Physics of the Earth and Planetary Interiors* 100: 61–79.

Karato S, Riedel MR, and Yuen DA (2001) Rheological structure and deformation of subducted slabs in the mantle transition zone: Implications for mantle circulation and deep earthquakes. *Physics of the Earth and Planetary Interiors* 127: 83–108.

Karato S, Wang ZC, Liu B, and Fujino K (1995) Plastic-deformation of garnets – Systematics and implications for the rheology of the mantle transition zone. *Earth and Planetary Science Letters* 130: 13–30.

Karato S and Wu P (1993) Rheology of the upper mantle – A synthesis. *Science* 260: 771–778.

Kasting JF, Eggler DH, and Raeburn SP (1993a) Mantle redox evolution and the oxidation state of the Archean atmosphere. *Journal of Geology* 101: 245–257.

Kasting JF, Whitmore DP, and Reynolds RT (1993b) Habitable zones around main sequence stars. *Icarus* 101: 108–128.

Katz RF and Rudge JF (2011) The energetics of melting fertile heterogeneities within the depleted mantle. *Geochemistry, Geophysics, Geosystems* 12: Q0AC16.

Kawai K, Tsuchiya T, Tsuchiya J, and Maruyama S (2009) Lost primordial continents. *Gondwana Research* 16: 581–586.

Keller T and Tackley PJ (2009) Towards self-consistent modelling of the Martian dichotomy: The influence of low-degree convection on crustal thickness distribution. *Icarus* 202: 429–443.

Kellogg LH (1997) Growing the Earth's D″ layer: Effect of density variations at the core-mantle boundary. *Geophysical Research Letters* 24: 2749–2752.

Kellogg JB (2004) *Towards and Understanding of Chemical and Isotopic Heterogeneity in the Earth's Mantle*. Cambridge, MA: Harvard University.

Kellogg LH, Hager BH, and van der Hilst RD (1999) Compositional stratification in the deep mantle. *Science* 283: 1881–1884.

Kellogg JB, Jacobsen SB, and O'Connell RJ (2002) Modeling the distribution of isotopic ratios in geochemical reservoirs. *Earth and Planetary Science Letters* 204: 183–202.

Kellogg JB, Jacobsen SB, and O'Connell RJ (2007) Modeling lead isotopic heterogeneity in mid-ocean ridge basalts. *Earth and Planetary Science Letters* 262: 328–342.

Kellogg LH and Stewart CA (1991) Mixing by chaotic convection in an infinite Prandtl number fluid and implications for mantle convection. *Physics of Fluids A: Fluid Dynamics* 3: 1374–1378.

Kellogg LH and Turcotte DL (1986) Homogenization of the mantle by convective mixing and diffusion. *Earth and Planetary Science Letters* 81: 371–378.

Kellogg LH and Turcotte DL (1990) Mixing and the distribution of heterogeneities in a chaotically convecting mantle. *Journal of Geophysical Research* 95: 421–432.

Kellogg LH and Wasserburg GJ (1990) The role of plumes in mantle helium fluxes. *Earth and Planetary Science Letters* 99: 276–289.

Kennett BLN and Gorbatov A (2004) Seismic heterogeneity in the mantle – Strong shear wave signature of slabs from joint tomography. *Physics of the Earth and Planetary Interiors* 146: 87–100.

Kennett BLN, Widiyantoro S, and Vanderhilst RD (1998) Joint seismic tomography for bulk sound and shear-wave speed in the earths mantle. *Journal of Geophysical Research, Solid Earth* 103: 12469–12493.

Kesson SE, Gerald JDF, and Shelley JM (1998) Mineralogy and dynamics of a pyrolite lower mantle. *Nature* 393: 252–255.

Klein EM, Langmuir CH, Zindler A, Staudigel H, and Hamelin H (1988) Isotope evidence of a mantle convection boundary at the Australian-Antarctic Discordance. *Nature* 333: 623–629.

Knittle E and Jeanloz R (1989) Simulating the core-mantle boundary – An experimental study of high-pressure reactions between silicates and liquid iron. *Geophysical Research Letters* 16: 609–612.

Knittle E and Jeanloz R (1991) Earths core-mantle boundary – Results of experiments at high-pressures and temperatures. *Science* 251: 1438–1443.

Kogiso T, Hirschmann MM, and Frost DJ (2003) High-pressure partial melting of garnet pyroxenite: Possible mafic lithologies in the source of ocean island basalts. *Earth and Planetary Science Letters* 216: 603–617.

Kogiso T, Hirschmann MM, and Reiners PW (2004) Length scales of mantle heterogeneities and their relationship to ocean island basalt geochemistry. *Geochimica et Cosmochimica Acta* 68: 345–360.

Komabayashi T, Maruyama S, and Rino S (2009) A speculation on the structure of the D″ layer: The growth of anti-crust at the core-mantle boundary through the subduction history of the Earth. *Gondwana Research* 15: 342–353.

Korenaga J (2003) Energetics of mantle convection and the fate of fossil heat. *Geophysical Research Letters* 30: 1437.
Kramers JD and Tolstikhin IN (1997) Two terrestrial lead isotope paradoxes, forward transport modelling, core formation and the history of the continental crust. *Chemical Geology* 139: 75–110.
Kumagai I, Davaille A, Kurita K, and Stutzmann E (2008) Mantle plumes: Thin, fat, successful, or failing? Constraints to explain hot spot volcanism through time and space. *Geophysical Research Letters* 35: L16301. http://dx.doi.org/10.1029/2008GL035079.
Kump LR, Kasting JF, and Barley ME (2001) Rise of atmospheric oxygen and the "upside-down" Archean mantle. *Geochemistry, Geophysics, Geosystems* 2: U1–U10. http://dx.doi.org/10.1029/2000GC000114.
Kurz MD, Jenkins WJ, Hart SR, and Clague D (1983) Helium isotopic variations in volcanic rocks from Loihi seamount and the island of Hawaii. *Earth and Planetary Science Letters* 66: 388–406.
Kyvalova H, Cadek O, and Yuen DA (1995) Correlation analysis between subduction in the last 180 Myr and lateral seismic structure of the lower mantle: Geodynamical implications. *Geophysical Research Letters* 22: 1281–1284.
Labrosse S (2003) Thermal and magnetic evolution of the Earth's core. *Physics of the Earth and Planetary Interiors* 140: 127–143.
Labrosse S, Hernlund JW, and Coltice N (2007) A crystallising dense magma ocean at the base of Earth's mantle. *Nature* 450: 866–869.
Labrosse S, Poirier JP, and Le Mouel JL (1997) On cooling of the Earth's core. *Physics of the Earth and Planetary Interiors* 99: 1–17.
Lassak TM, McNamara AK, Garnero EJ, and Zhong S (2010) Core-mantle boundary topography as a possible constraint on lower mantle chemistry and dynamics. *Earth and Planetary Science Letters* 289: 232–241.
Lassak TM, McNamara AK, and Zhong S (2007) Influence of thermochemical piles on topography at Earth's core-mantle boundary. *Earth and Planetary Science Letters* 261: 443–455.
Lassiter JC (2004) Role of recycled oceanic crust in the potassium and argon budget of the Earth: Toward a resolution of the "missing argon" problem. *Geochemistry, Geophysics, Geosystems* 5: Q11012, http://dx.doi.org/11010.11029/12004GC000711.
Lay T and Garnero EJ (2004) Core-mantle boundary structures and processes. In: Sparks RSJ and Hawkesworth CJ (eds.) *The State of the Planet: Frontiers and Challenges in Geophysics*. Washington, DC: AGU, http://dx.doi.org/10.1029/1150GM1004.
Lay T, Garnero EJ, and Williams Q (2004) Partial melting in a thermo-chemical boundary layer at the base of the mantle. *Physics of the Earth and Planetary Interiors* 146: 441–467.
Le Bars M and Davaille A (2002) Stability of thermal convection in two superimposed miscible viscous fluids. *Journal of Fluid Mechanics* 471: 339–363.
Le Bars M and Davaille A (2004a) Large interface deformation in two-layer thermal convection of miscible viscous fluids. *Journal of Fluid Mechanics* 499: 75–110.
Le Bars M and Davaille A (2004b) Whole layer convection in a heterogeneous planetary mantle. *Journal of Geophysical Research* 109: B03403, http://dx.doi.org/03410.01029/02003JB002617.
Leahy GM and Bercovici D (2004) The influence of the transition zone water filter on convective circulation in the mantle. *Geophysical Research Letters* 31. http://dx.doi.org/10.1029/2004GL021206.
Leahy GM and Bercovici D (2007) On the dynamics of a hydrous melt layer above the transition zone. *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2006JB004631.
Leahy GM and Bercovici D (2010) Reactive infiltration of hydrous melt above the mantle transition zone. *Journal of Geophysical Research* 115: B08406.
Lecuyer C and Ricard Y (1999) Long-term fluxes and budget of ferric iron: Implication for the redox states of the Earth's mantle and atmosphere. *Earth and Planetary Science Letters* 165: 197–211.
Lee CA, Luffi P, and Chin EJ (2011) Building and destroying continental mantle. *Annual Review of Earth and Planetary Sciences* 39: 59–90.
Lee C-T, Luffi P, Hoink T, Li J, Dasgupta R, and Hernlund J (2010) Upside-down differentiation and generation of a 'primordial' lower mantle. *Nature* 463: 930–933.
Lee C-TA, Shen B, Slotnick BS, et al. (2013) Continental arc-island arc fluctuations, growth of crustal carbonates, and long-term climate change. *Geosphere* 9: 21–36.
Lenardic A and Kaula WM (1993) A numerical treatment of geodynamic viscous-flow problems involving the advection of material interfaces. *Journal of Geophysical Research: Solid Earth* 98: 8243–8260.
Li M and McNamara AK (2013) The difficulty for subducted oceanic crust to accumulate at the Earth's core-mantle boundary. *Journal of Geophysical Research, Solid Earth* 118: 1807–1816.
Lin S-C and Van Keken PE (2006a) Dynamics of thermochemical plumes: 1. Plume formation and entrainment of a dense layer. *Geochemistry, Geophysics, Geosystems* 7. http://dx.doi.org/10.1029/2005GC001071.
Lin S-C and Van Keken PE (2006b) Dynamics of thermochemical plumes: 2. Complexity of plume structures and its implications for mapping of mantle plumes. *Geochemistry, Geophysics, Geosystems* 7. http://dx.doi.org/10.1029/2005GC001072.
Lithgow-Bertelloni C, Richards MA, Ricard Y, O'Connell RJ, and Engebretson DC (1993) Toroidal-poloidal partitioning of plate motions since 120 Ma. *Geophysical Research Letters* 20: 375–378.
Luo S-N, Ni S, and Helmberger DV (2001) Evidence for a sharp lateral variation of velocity at the core-mantle boundary from multipathed PKPab. *Earth and Planetary Science Letters* 189: 155–164.
Lyubetskaya T and Korenaga J (2007a) Chemical composition of Earth's primitive mantle and its variance, 1, Method and results. *Journal of Geophysical Research* 112. http://dx.doi.org/10.1029/2005JB004223.
Lyubetskaya T and Korenaga J (2007b) Chemical composition of Earth's primitive mantle and its variance, 2, Implications for global geodynamics. *Journal of Geophysical Research* 112(B3). http://dx.doi.org/10.1029/2005JB004224.
Machetel P and Weber P (1991) Intermittent layered convection in a model mantle with an endothermic phase-change at 670 km. *Nature* 350: 55–57.
Malvern LE (1969) *Introduction to the Mechanics of a Continuous Medium*. Englewood Cliffs, NJ: Prentice Hall.
Mambole A and Fleitout L (2002) Petrological layering induced by an endothermic phase transition in the Earth's mantle – Art. no. 2044. *Geophysical Research Letters* 29: 2044.
Manga M (1996) Mixing of heterogeneities in the mantle – Effect of viscosity differences. *Geophysical Research Letters* 23: 403–406.
Manga M and Jeanloz R (1996) Implications of a metal-bearing chemical-boundary layer in D″ for mantle dynamics. *Geophysical Research Letters* 23: 3091–3094.
Mao WL, Mao HK, Sturhahn W, et al. (2006) Iron-rich postperovskite and the origin of ultralow-velocity zones. *Science* 312: 564–565.
Martin H (1986) Effect of steeper archean geothermal gradient on geochemistry of subduction zone magmas. *Geology* 14: 753–756.
Masters G, Laske G, Bolton H, and Dziewonski A (2000) The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: Implications for chemical and thermal structure. In: Karato S, Forte AM, Liebermann RC, Masters G, and Stixrude L (eds.) *Geophysical Monograph on Mineral Physics and Seismic Tomography from the Atomic to the Global Scale*, pp. 63–87. Washington, DC: American Geophysical Union.
Matsuda JMS, Ozima M, Ito K, Ohtaka O, and Ito E (1993) Noble gas partitioning between metal and silicate under high pressures. *Science* 259: 788–790.
McDonough WF and Sun S-S (1995) The composition of the Earth. *Chemical Geology* 120: 223–253.
McGovern PJ and Schubert G (1989) Thermal evolution of the Earth – Effects of volatile exchange between atmosphere and interior. *Earth and Planetary Science Letters* 96: 27–37.
McKenzie D (1979) Finite deformation during fluid flow. *Geophysical Journal of the Royal Astronomical Society* 58: 689–715.
McKenzie D and Onions RK (1992) Partial melt distributions from inversion of rare-earth element concentrations (J Petrol, Vol 32, Pg 1021, 1991). *Journal of Petrology* 33: 1453.
McNamara AK, Garnero EJ, and Rost S (2010) Tracking deep mantle reservoirs with ultra-low velocity zones. *Earth and Planetary Science Letters* 299: 1–9.
McNamara AK and Van Keken PE (2000) Cooling of the Earth: A parameterized convection study of whole versus layered models. *Geochemistry, Geophysics, Geosystems* 1: 1027. http://dx.doi.org/10.1029/2000GC000045.
McNamara AK and Zhong S (2004a) The influence of thermochemical convection on the fixity of mantle plumes. *Earth and Planetary Science Letters* 222: 485–500.
McNamara AK and Zhong S (2004b) Thermochemical structures within a spherical mantle: Superplumes or piles? *Journal of Geophysical Research* 109: B07402. http://dx.doi.org/10.1029/2003JB00247.
McNamara AK and Zhong S (2005) Thermochemical piles beneath Africa and the Pacific Ocean. *Nature* 437: 1136–1139.
Megnin C, Bunge HP, Romanowicz B, and Richards MA (1997) Imaging 3-D spherical convection models – What can seismic tomography tell us about mantle dynamics. *Geophysical Research Letters* 24: 1299–1302.
Meibom A and Anderson DL (2004) The statistical upper mantle assemblage. *Earth and Planetary Science Letters* 217: 123–139.
Meibom A, Anderson DL, Sleep NH, et al. (2003) Are high He-3/He-4 ratios in oceanic basalts an indicator of deep-mantle plume components? Earth Planet. *Science Letters* 208: 197–204.

Meibom A, Sleep NH, Chamberlain CP, et al. (2002) Re-Os isotopic evidence for long-lived heterogeneity and equilibration processes in the Earth's upper mantle. *Nature* 419: 705–708.

Meibom A, Sleep NH, Zahnle K, and Anderson DL (2005) Models for noble gases in mantle geochemistry: Some observations and alternatives. In: *Plumes, Plates, and Paradigms*, pp. 347–363, Geological Society of America Special Paper 388.

Melosh HJ (1990) Giant impacts and the thermal state of the early Earth. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*. New York: Oxford University Press.

Metcalfe G, Bina CR, and Ottino JM (1995) Kinematic considerations for mantle mixing. *Geophysical Research Letters* 22: 743–746.

Michard A and Albarède F (1985) Hydrothermal uranium uptake at ridge crests. *Nature* 317: 61–88.

Montague NL and Kellogg LH (2000) Numerical models for a dense layer at the base of the mantle and implications for the geodynamics of D″. *Journal of Geophysical Research* 105: 11101–11114.

Montague NL, Kellogg LH, and Manga M (1998) High Rayleigh number thermo-chemical models of a dense boundary layer in D″. *Geophysical Research Letters* 25: 2345–2348.

Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl ER, and Hung SH (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303: 338–343.

Moore WB and Webb AAG (2013) Heat-pipe Earth. *Nature* 501: 501–505.

Moreira M, Breddam K, Curtice J, and Kurz MD (2001) Solar neon in the Icelandic mantle: New evidence for an undegassed lower mantle. *Earth and Planetary Science Letters* 185: 15–23.

Moresi L and Solomatov V (1998) Mantle convection with a brittle lithosphere – Thoughts on the global tectonic styles of the Earth and Venus. *Geophysical Journal International* 133: 669–682.

Mosca I, Cobden L, Deuss A, Ritsema J, and Trampert J (2012) Seismic and mineralogical structures of the lower mantle from probabilistic tomography. *Journal of Geophysical Research* 117: B06304.

Mosenfelder JL, Asimow PD, and Ahrens TJ (2007) Thermodynamic properties of Mg_2SiO_4 liquid at ultra-high pressures from shock measurements to 200 GPa on forsterite and wadsleyite. *Journal of Geophysical Research* 112: B06208, http://dx.doi.org/06210.01029/02006jb004364.

Mosenfelder JL, Asimow PD, Frost DJ, et al. (2009) The $MgSiO_3$ system at high pressure: Thermodynamic properties of perovskite, postperovskite, and melt from global inversion of shock and static compression data. *Journal of Geophysical Research* 114: B01203, http://dx.doi.org/01210.01029/02008jb005900.

Muller R (1992) The performance of classical versus modern finite-volume advection schemes for atmospheric modeling in a one-dimensional test-bed. *The Monthly Weather Review* 120: 1407–1415.

Murakami M, Hirose K, Kawamura K, Sata N, and Ohishi Y (2004) Post-perovskite phase transition in $MgSiO_3$. *Science* 304: 855–858.

Murakami M, Hirose K, Sata N, and Ohishi Y (2005) Post-perovskite phase transition and mineral chemistry in the pyrolitic lowermost mantle. *Geophysical Research Letters* 32. http://dx.doi.org/10.1029/2004GL021956.

Murphy DT, Kamber BS, and Collerson KD (2003) A refined solution to the first terrestrial Pb-isotope paradox. *Journal of Petrology* 44: 39–53.

Murthy VR and Karato S (1997) Core formation and chemical-equilibrium in the earth. 2. Chemical consequences for the mantle and core. *Physics of the Earth and Planetary Interiors* 100: 81–95.

Nakagawa T and Buffett BA (2005) Mass transport mechanism between the upper and lower mantle in numerical simulations of thermochemical mantle convection with multicomponent phase changes. *Earth and Planetary Science Letters* 230: 11–27.

Nakagawa T and Tackley PJ (2004a) Effects of a perovskite-post perovskite phase change near the core-mantle boundary on compressible mantle convection. *Geophysical Research Letters* 31: L16611. http://dx.doi.org/10.1029/2004GL020648.

Nakagawa T and Tackley PJ (2004b) Effects of thermo-chemical mantle convection on the thermal evolution of the Earth's core. *Earth and Planetary Science Letters* 220: 107–119.

Nakagawa T and Tackley PJ (2005a) Deep mantle heat flow and thermal evolution of Earth's core in thermo-chemical multiphase models of mantle convection. *Geochemistry, Geophysics, Geosystems* 6: Q08003. http://dx.doi.org/10.1029/2005GC000967.

Nakagawa T and Tackley PJ (2005b) The interaction between the post-perovskite phase change and a thermo-chemical boundary layer near the core-mantle boundary. *Earth and Planetary Science Letters* 238: 204–216.

Nakagawa T and Tackley PJ (2005c) Three-dimensional numerical simulations of thermo-chemical multiphase convection in Earth's mantle. In: *Proceedings of the Third MIT Conference on Computational Fluid and Solid Mechanics*.

Nakagawa T and Tackley PJ (2006) Three-dimensional structures and dynamics in the deep mantle: Effects of post-perovskite phase change and deep mantle layering. *Geophysical Research Letters* 33: L12S11. http://dx.doi.org/10.1029/2006GL025719.

Nakagawa T and Tackley PJ (2010) Influence of initial CMB temperature and other parameters on the thermal evolution of Earth's core resulting from thermo-chemical spherical mantle convection. *Geochemistry, Geophysics, Geosystems* 11: Q06001. http://dx.doi.org/10.1029/2010GC003031.

Nakagawa T and Tackley PJ (2011) Effects of low-viscosity post-perovskite on thermo-chemical convection in a 3-D spherical shell. *Geophysical Research Letters* 38: L04309. http://dx.doi.org/10.1029/2010GL046494.

Nakagawa T and Tackley PJ (2012) Influence of magmatism on mantle cooling, surface heat flow and Urey ratio. *Earth and Planetary Science Letters* 329–330: 1–10.

Nakagawa T and Tackley PJ (2013) Implications of high core thermal conductivity on Earth's coupled mantle and core evolution. *Geophysical Research Letters* 40: 2652–2656.

Nakagawa T and Tackley PJ (2014) Influence of combined primordial layering and recycled MORB on the coupled thermal evolution of Earth's mantle and core. *Geochemistry, Geophysics, Geosystems* 15: 619–633. http://dx.doi.org/10.1002/2013GC005128.

Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2009) Incorporating self-consistently calculated mineral physics into thermo-chemical mantle convection simulations in a 3D spherical shell and its influence on seismic anomalies in Earth's mantle. *Geochemistry, Geophysics, Geosystems* 10. http://dx.doi.org/10.1029/2008GC002280.

Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2010) The influence of MORB and harzburgite composition on thermo-chemical mantle convection in a 3-D spherical shell with self-consistently calculated mineral physics. *Earth and Planetary Science Letters* 296: 403–412.

Nakagawa T, Tackley PJ, Deschamps F, and Connolly JAD (2012) Radial 1-D seismic structures in the deep mantle in mantle convection simulations with self-consistently calculated mineralogy. *Geochemistry, Geophysics, Geosystems* 13: Q11002. http://dx.doi.org/10.1029/2012GC004325.

Namiki A (2003) Can the mantle entrain D″? *Journal of Geophysical Research* 108: 2487. http://dx.doi.org/10.1029/2002JB002315.

Newsom HE, White WM, Jochum KP, and Hofmann AW (1986) Siderophile and chalcophile element abundances in oceanic basalts. *Earth and Planetary Science Letters* 80: 299–313.

Ni SV, Helmberger D, and Tromp J (2005) Three-dimensional structure of the African superplume from waveform modelling. *Geophysical Journal International* 161: 283–294.

Ni S, Tan E, Gurnis M, and Helmberger DV (2002) Sharp sides to the African superplume. *Science* 296: 1850–1852.

Nimmo F, Price GD, Brodholt J, and Gubbins D (2004) The influence of potassium on core and geodynamo evolution. *Geophysical Journal International* 156: 363–376.

Nomura R, Ozawa H, Tateno S, et al. (2011) Spin crossover and iron-rich silicate melt in the Earth/'s deep mantle. *Nature* 473: 199–202.

Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of $MgSiO_3$ in Earth's D″ layer. *Nature* 430: 445–448.

Ogawa M (1988) Numerical experiments on coupled magmatism-mantle convection system – Implications for mantle evolution and archean continental crusts. *Journal of Geophysical Research – B: Solid Earth and Planets* 93: 15119–15134.

Ogawa M (1993) A numerical-model of a coupled magmatism mantle convection system in venus and the Earths mantle beneath archean continental crusts. *Icarus* 102: 40–61.

Ogawa M (1994) Effects of chemical fractionation of heat-producing elements on mantle evolution inferred from a numerical-model of coupled magmatism mantle convection system. *Physics of the Earth and Planetary Interiors* 83: 101–127.

Ogawa M (1997) A bifurcation in the coupled magmatism-mantle convection system and its implications for the evolution of the Earth's upper mantle. *Physics of the Earth and Planetary Interiors* 102: 259–276.

Ogawa M (2000a) Coupled magmatism-mantle convection system with variable viscosity. *Tectonophysics* 322: 1–18.

Ogawa M (2000b) Numerical models of magmatism in convecting mantle with temperature-dependent viscosity and their implications for Venus and Earth. *Journal of Geophysical Research* 105: 6997–7012.

Ogawa M (2003) Chemical stratification in a two-dimensional convecting mantle with magmatism and moving plates. *Journal of Geophysical Research* 108: 2561. http://dx.doi.org/10.1029/2002JB002205.

Ogawa M (2007) Superplumes, plates, and mantle magmatism in two-dimensional numerical models. *Journal of Geophysical Research* 112: B06404.

Ogawa M (2010) Variety of plumes and the fate of subducted basaltic crusts. *Physics of the Earth and Planetary Interiors* 183: 366–375.

Ogawa M (2013) Two-stage evolution of the Earth's mantle inferred from numerical simulations of coupled magmatism-mantle convection system with tectonic plates. *Journal of Geophysical Research* 119: 2462–2486.
Ogawa M and Nakamura H (1998) Thermochemical regime of the early mantle inferred from numerical models of the coupled magmatism-mantle convection system with the solid-solid phase transitions at depths around 660 km. *Journal of Geophysical Research* 103: 12161–12180.
Ogawa M and Yanagisawa T (2011) Numerical models of Martian mantle evolution induced by magmatism and solid-state convection beneath stagnant lithosphere. *Journal of Geophysical Research* 116: E08008. http://dx.doi.org/10.1029/2010JE003777.
Oldham D and Davies JH (2004) Numerical investigation of layered convection in a three-dimensional shell with application to planetary mantles. *Geochemistry, Geophysics, Geosystems* 5: Q12C04. http://dx.doi.org/10.1029/2003GC000603.
Olson P (1984) An experimental approach to thermal-convection in a 2-layered mantle. *Journal of Geophysical Research* 89: 1293–1301.
Olson P and Kincaid C (1991) Experiments on the interaction of thermal convection and compositional layering at the base of the mantle. *Journal of Geophysical Research* 96: 4347–4354.
Olson P, Yuen DA, and Balsiger D (1984a) Convective mixing and the fine-structure of mantle heterogeneity. *Physics of the Earth and Planetary Interiors* 36: 291–304.
Olson P, Yuen DA, and Balsiger D (1984b) Mixing of passive heterogeneities by mantle convection. *Journal of Geophysical Research* 89: 425–436.
O'Neill HS and Palme H (2008) Collisional erosion and the non-chondritic composition of the terrestrial planets. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 366: 4205–4238.
O'Nions RK, Evensen NM, and Hamilton PJ (1979) Geochemical modeling of mantle differentiation and crustal growth. *Journal of Geophysical Research* 84: 6091–6101.
O'Nions RK and Oxburgh ER (1983) Heat and helium in the Earth. *Nature* 306: 429.
O'Nions RK and Tolstikhin IN (1996) Limits on the mass flux between lower and upper mantle and stability of layering. *Earth and Planetary Science Letters* 139: 213–222.
Ono S, Ito E, and Katsura T (2001) Mineralogy of subducted basaltic crust (MORB) from 25 to 37 GPa, and chemical heterogeneity of the lower mantle. *Earth and Planetary Science Letters* 190: 57–63.
Ono S, Oganov AR, and Ohishi Y (2005a) In situ observations of phase transition between perovskite and $CaIrO_3$-type phase in $MgSiO_3$ and pyrolitic mantle composition. *Earth and Planetary Science Letters* 236: 914–932.
Ono S, Oshishi Y, Isshiki M, and Watanuki T (2005b) In situ X-ray observations of phase assemblages in peridotite and basalt compositions at lower mantle conditions: Implications for density of subducted oceanic plate. *Journal of Geophysical Research* 110: B02208. http://dx.doi.org/10.1029/2004JB0003196.
O'Reilly TC and Davies GF (1981) Magma transport of heat on Io: A mechanism allowing a thick lithosphere. *Geophysical Research Letters* 8: 313–316.
Otsuka K and Karato S (2012) Deep penetration of molten iron into the mantle caused by a morphological instability. *Nature* 492: 243–246.
Ottino JM (1989) *The Kinematics of Mixing: Stretching, Chaos, and Transport*. Cambridge: Cambridge University Press.
Ozima M and Kudo K (1972) Excess argon in submarine basalts and an Earth-atmosphere evolution model. *Nature* 239: 23–24.
Palme H and O'Neill HSC (2003) Cosmochemical estimates of mantle composition. In: Carlson RW (ed.) *Treatise on Geochemistry*, pp. 1–38. Oxford: Elsevier.
Parkinson IJ, Hawkesworth CJ, and Cohen AS (1998) Ancient mantle in a modern arc: Osmium isotopes in Izu-Bonin-Mariana forearc peridotites. *Science* 281: 2011–2013.
Parman SW, Kurz MD, Hart SR, and Grove TL (2005) Helium solubility in olivine and implications for high $^3He/^4He$ in ocean island basalts. *Nature* 437: 1140–1143.
Pearson DG, Davies GR, and Nixon PH (1993) Geochemical constraints on the petrogenesis of diamond facies pyroxenites from the Beni Bousera peridotite massif, north Morocco. *Journal of Petrology* 34: 125–172.
Peltier WR and Solheim LP (1992) Mantle phase-transitions and layered chaotic convection. *Geophysical Research Letters* 19: 321–324.
Pertermann M (2003) Partial melting experiments on a MORB-like pyroxenite between 2 and 3 GPa: Constraints on the presence of pyroxenite in basalt source regions from solidus location and melting rate. *Journal of Geophysical Research* 108: 2125.
Petford N, Yuen D, Rushmer T, Brodholt J, and Stackhouse S (2005) Shear-induced material transfer across the core-mantle boundary aided by the post-perovskite phase transition. *Earth, Planets and Space* 57: 459–464.
Phipps Morgan J (1998) Thermal and rare gas evolution of the mantle. *Chemical Geology* 145: 431–445.
Phipps Morgan J (1999) Isotope topology of individual hotspot basalt arrays: Mixing curves or melt extraction trajectories? *Geochemistry, Geophysics, Geosystems* 1: 1003. http://dx.doi.org/10.1029/1999GC000004.
Phipps Morgan J (2001) Thermodynamics of pressure release melting of a veined plum pudding mantle. *Geochemistry, Geophysics, Geosystems* 2, Paper number 2000GC000049 [000015,000429 words, 000010 figures, 000041 table, 000042 appendix figures, 000041 appendix table].
Phipps Morgan J and Morgan J (1999) Two-stage melting and the geochemical evolution of the mantle: A recipe for mantle plum-pudding. *Earth and Planetary Science Letters* 170: 215–239.
Poirier JP (1993) Core-infiltrated mantle and the nature of the D″ layer. *Journal of Geomagnetism and Geoelectricity* 45: 1221–1227.
Poirier JP and Le Mouel JL (1992) Does infiltration of core material into the lower mantle affect the observed geomagnetic field? *Physics of the Earth and Planetary Interiors* 73: 29–37.
Poirier JP, Malavergne V, and Le Mouel JL (1998) Is there a thin electrically conducting layer at the base of the mantle? In: Gurnis M, Wysession ME, Knittle E, and Buffett BA (eds.) *The Core-Mantle Boundary Region*, pp. 131–137. Washington, DC: American Geophysical Union.
Polve M and Allegre CJ (1980) Orogenic lherzolite complexes studies by 87Rb-87Sr: A clue to understand the mantle convection processes? Earth Planet. *Science Letters* 51: 71–93.
Porcelli D and Halliday AN (2001) The core as a possible source of mantle helium. *Earth and Planetary Science Letters* 192: 45–56.
Porcelli D and Wasserburg GJ (1995) Mass transfer of helium, neon, argon and xenon through a steady-state upper mantle. *Geochimica et Cosmochimica Acta* 23: 4921–4937.
Pozzo M, Davies C, Gubbins D, and Alfe D (2012) Thermal and electrical conductivity of iron at Earth's core conditions. *Nature* 485: 355–358.
Puster P and Jordan TH (1994) Stochastic-analysis of mantle convection experiments using 2-point correlation-functions. *Geophysical Research Letters* 21: 305–308.
Puster P and Jordan TH (1997) How stratified is mantle convection? *Journal of Geophysical Research* 102: 7625–7646.
Puster P, Jordan TH, and Hager BH (1995) Characterization of mantle convection experiments using 2-point correlation-functions. *Journal of Geophysical Research, Solid Earth* 100: 6351–6365.
Rapp RP, Irifune T, Shimizu N, Nishiyama N, Norman MD, and Inoue T (2008) Subduction recycling of continental sediments and the origin of geochemically enriched reservoirs in the deep mantle. *Earth and Planetary Science Letters* 271: 14–23.
Reisberg I and Zindler A (1986) Extreme isotopic variability in the upper mantle: Evidence from Ronda. *Earth and Planetary Science Letters* 81: 29–45.
Revenaugh J and Meyer R (1997) Seismic evidence of partial melt within a possibly ubiquitous low-velocity layer at the base of the mantle. *Science* 277: 670–673.
Revenaugh J and Sipkin SA (1994) Seismic evidence for silicate melt atop the 410-km mantle discontinuity. *Nature* 369: 474–476.
Ricard Y, Froidevaux C, and Fleitout L (1988) Global plate motion and the geoid – A physical model. *Geophysical Journal International* 93: 477–484.
Ricard Y, Vigny C, and Froidevaux C (1989) Mantle heterogeneities, geoid, and plate motion – A Monte-Carlo inversion. *Journal of Geophysical Research* 94: 13739–13754.
Richard GC and Bercovici D (2009) Water-induced convection in the Earth's mantle transition zone. *Journal of Geophysical Research* 114: B01205. http://dx.doi.org/10.1029/2008JB005734.
Richard G, Bercovici D, and Karato S-I (2006) Slab dehydration in the Earth's mantle transition zone. *Earth and Planetary Science Letters* 251: 156–167.
Richard G, Monnereau M, and Ingrin J (2002) Is the transition zone an empty water reservoir? Inferences from numerical models of mantle dynamics. *Earth and Planetary Science Letters* 205: 37–51.
Richards MA and Davies GF (1989) On the separation of relatively buoyant components from subducted lithosphere. *Geophysical Research Letters* 16: 831–834.
Richards MA and Engebretson DC (1992) Large-scale mantle convection and the history of subduction. *Nature* 355: 437–440.
Richards MA and Griffiths RW (1989) Thermal entrainment by deflected mantle plumes. *Nature* 342: 900–902.
Richardson LF (1926) Atmospheric diffusion shown on a distance-neighbor graph. *Proceedings of the Royal Society of London* 110: 709–737.
Richter FM, Daly SF, and Nataf HC (1982) A parameterized model for the evolution of isotopic heterogeneities in a convecting system. *Earth and Planetary Science Letters* 60: 178–194.
Richter FM and Johnson CE (1974) Stability of a chemically layered mantle. *Journal of Geophysical Research* 79: 1635–1639.
Richter FM and McKenzie D (1981) On some consequences and possible causes of layered mantle convection. *Journal of Geophysical Research* 86: 6133–6142.

Richter FM and Ribe NM (1979) On the importance of advection in determining the local isotopic composition of the mantle. *Earth and Planetary Science Letters* 43: 212–222.

Ricolleau A, Perrillat J-P, Fiquet G, et al. (2010) Phase relations and equation of state of a natural MORB: Implications for the density profile of subducted oceanic crust in the Earth's lower mantle. *Journal of Geophysical Research* 115: B08202. http://dx.doi.org/08210.01029/02009jb006709.

Righter K and Drake MJ (2003) Partition coefficients at high pressure and temperature. In: Carlson CW (ed.) *Treatise on Geochemistry*. *The Mantle and Core*, vol. 2, pp. 425–449. Amsterdam: Elsevier.

Ringwood AE (1990a) Earliest history of the Earth-Moon system. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 101–134. New York: Oxford University Press.

Ringwood AE (1990b) Slab-mantle interactions. 3. Petrogenesis of intraplate magmas and structure of the upper mantle. *Chemical Geology* 82: 187–207.

Ringwood AE (1991) Phase-transformations and their bearing on the constitution and dynamics of the mantle. *Geochimica et Cosmochimica Acta* 55: 2083–2110.

Ringwood AE (1994) Role of the transition zone and 660 km discontinuity in mantle dynamics. *Physics of the Earth and Planetary Interiors* 86: 5–24.

Ringwood AE and Irifune T (1988) Nature of the 650-km seismic discontinuity – Implications for mantle dynamics and differentiation. *Nature* 331: 131–136.

Rizo H, Boyet M, Blichert-Toft J, O'Neil J, Rosing MT, and Paquette JL (2012) The elusive Hadean enriched reservoir revealed by ^{142}Nd deficits in Isua Archaean rocks. *Nature* 491: 96–100.

Rohrbach A and Schmidt MW (2011) Redox freezing and melting in the Earth's deep mantle resulting from carbon-iron redox coupling. *Nature* 472: 209–212.

Romanowicz B (2001) Can we resolve 3D density heterogeneity in the lower mantle? *Geophysical Research Letters* 28: 1107–1110.

Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal-silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.

Rudge JF (2006) Mantle pseudo-isochrons revisited. *Earth and Planetary Science Letters* 249: 494–513.

Rudge JF, Bercovici D, and Spiegelman M (2011) Disequilibrium melting of a two phase multicomponent mantle. *Geophysical Journal International* 184: 699–718.

Rudge JF, McKenzie D, and Haynes PH (2005) A theoretical approach to understanding the isotopic heterogeneity of mid-ocean ridge basalt. *Geochimica et Cosmochimica Acta* 69: 3873–3887.

Rudnick RL (1995) Making continental crust. *Nature (London)* 378: 571–578.

Rudnick RL and Fountain DM (1995) Nature and composition of the continental crust; a lower crustal perspective. *Reviews of Geophysics* 33: 267–309.

Ruepke L, Phipps Morgan J, Hort M, and Connolly JAD (2004) Serpentine and the subduction zone water cycle. *Earth and Planetary Science Letters* 223: 17–34.

Saal AE, Hart SR, Shimizu N, Hauri EH, and Layne GD (1998) Pb isotopic variability in melt inclusions from oceanic island basalts, Polynesia. *Science* 282: 1481–1484.

Saal AE, Hauri EH, Langmuir CH, and Perfit MR (2002) Vapour undersaturation in primitive mid-ocean-ridge basalt and the volatile content of Earth's upper mantle. *Nature* 419: 451–455.

Saltzer RL, Stutzmann E, and van der Hilst RD (2004) Poisson's ratio in the lower mantle beneath Alaska: Evidence for compositional heterogeneity. *Journal of Geophysical Research* 109: B06301, http://dx.doi.org/06310.01029/02003jb002712.

Samuel H and Bercovici D (2006) Oscillating and stagnating plumes in the Earth's lower mantle. *Earth and Planetary Science Letters* 248: 90–105.

Samuel H and Farnetani CG (2003) Thermochemical convection and helium concentrations in mantle plumes. *Earth and Planetary Science Letters* 207: 39–56.

Samuel H and Tackley PJ (2008) Dynamics of core formation and equilibrium by negative diapirism. *Geochemistry, Geophysics, Geosystems* 9. http://dx.doi.org/10.1029/2007GC001896.

Sandu C, Lenardic A, and McGovern P (2011) The effects of deep water cycling on planetary thermal evolution. *Journal of Geophysical Research, Solid Earth* 116: B12404.

Sarda P, Moreira M, and Staudacher T (1999) argon-Lead isotopic correlation in mid-Atlantic ridge basalts. *Science* 283: 666–668.

Schersten A, Elliott T, Hawkesworth C, and Norman M (2004) Tungsten isotope evidence that mantle plumes contain no contribution from the Earth's core. *Nature* 427: 234–237.

Schilling J-G (1973) Iceland mantle plume: Geochemical study of Reykjanes ridge. *Nature* 242: 565–571.

Schmalzl J and Hansen U (1994) Mixing the Earths mantle by thermal-convection – A scale-dependent phenomenon. *Geophysical Research Letters* 21: 987–990.

Schmalzl J, Houseman GA, and Hansen U (1995) Mixing properties of three-dimensional (3-D) stationary convection. *Physics of Fluids* 7: 1027–1033.

Schmalzl J, Houseman GA, and Hansen U (1996) Mixing in vigorous, time-dependent three-dimensional convection and application to Earth's mantle. *Journal of Geophysical Research* 101: 21847–21858.

Schmalzl J and Loddoch A (2003) Using subdivision surfaces and adaptive surface simplification algorithms for modeling chemical heterogeneities in geophysical flows. *Geochemistry, Geophysics, Geosystems* 4: 8303. http://dx.doi.org/10.1029/2003GC000578.

Schmandt B, Dueker K, Humphreys E, and Hansen S (2012) Hot mantle upwelling across the 660 beneath Yellowstone. *Earth and Planetary Science Letters* 331–332: 224–236.

Schott B and Yuen DA (2004) Influences of dissipation and rheology on mantle plumes coming from the D″-layer. *Physics of the Earth and Planetary Interiors* 146: 139–145.

Schott B, Yuen DA, and Braun A (2002) The influences of composition-and temperature- dependent rheology in thermal-chemical convection on entrainment of the D″-layer. *Physics of the Earth and Planetary Interiors* 129: 43–65.

Schubert G, Masters G, Olson P, and Tackley P (2004) Superplumes or plume clusters? *Physics of the Earth and Planetary Interiors* 146: 147–162.

Scrivner C and Anderson DL (1992) The effect of post pangea subduction on global mantle tomography and convection. *Geophysical Research Letters* 19: 1053–1056.

Shaw DM (1970) Trace element fractionation during anatexis. *Geochimica et Cosmochimica Acta* 34: 237–242.

Shaw AM, Hauri EH, Behn MD, Hilton DR, Macpherson CG, and Sinton JM (2012) Long-term preservation of slab signatures in the mantle inferred from hydrogen isotopes. *Nature Geoscience* 5: 224–228.

Shen Y, Solomon SC, Bjarnason IT, and Wolfe CJ (1998) Seismic evidence for a lower-mantle origin of the Iceland Plume. *Nature (London)* 395: 62–65.

Simmons NA, Forte AM, Boschi L, and Grand SP (2010) GyPSuM: A joint tomographic model of mantle density and seismic wave speeds. *Journal of Geophysical Research* 115: B12310.

Sleep NH (1984) Tapping of magmas from ubiquitous mantle heterogeneities – An alternative to mantle plumes. *Journal of Geophysical Research* 89: 10029–10041.

Sleep NH (1988) Gradual entrainment of a chemical layer at the base of the mantle by overlying convection. *Geophysical Journal* 95: 437–447.

Sleep NH (1990) Hotspots and mantle plumes – Some phenomenology. *Journal of Geophysical Research – B: Solid Earth and Planets* 95: 6715–6736.

Sleep NH (2000) Evolution of the mode of convection within terrestrial planets. *Journal of Geophysical Research* 105: 17563–17578.

Sleep NH and Zahnle K (2001) Carbon dioxide cycling and implications for climate on ancient Earth. *Journal of Geophysical Research* 106: 1373–1399.

Smith RB, Jordan M, Steinberger B, et al. (2009) Geodynamics of the Yellowstone hotspot and mantle plume: Seismic and GPS imaging, kinematics, and mantle flow. *Journal of Volcanology and Geothermal Research* 188: 26–56.

Smolarkiewicz PK and Margolin LG (1998) MPDATA: A finite-difference solver for geophysical flows. *Journal of Computational Physics* 140: 459–480.

Sneeringer M, Hart SR, and Shimizu N (1984) Strontium and samarium diffusion in diopside. *Geochimica et Cosmochimica Acta* 48: 1589–1608.

Sobolev AV (1996) Melt inclusions in minerals as a source of principal petrological information. *Petrology* 4: 209–220.

Sobolev AV, Hofmann AW, Kuzmin DV, et al. (2007) The amount of recycled crust in sources of mantle-derived melts. *Science* 316: 412–417.

Sobolev AV, Hofmann AW, and Nikogosian IK (2000) Recycled oceanic crust observed in 'ghost plagioclase' within the source of Mauna Loa lavas. *Nature* 404: 986–990.

Sobolev AV, Hofmann AW, Sobolev SV, and Nikogosian IK (2005) An olivine-free mantle source of Hawaiian shield basalts. *Science* 434: 590–597.

Sobolev AV and Shimizu N (1993) Ultra-depleted primary melt included in an olivine from the Mid-Atlantic Ridge. *Nature* 363: 151–154.

Solomatov VS (2000) Fluid dynamics of a terrestrial magma ocean. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*. Tucson, AZ: The University of Arizona Press.

Solomatov VS and Moresi LN (2002) Small-scale convection in the D″ layer. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2000JB000063.

Solomatov VS, Olson P, and Stevenson DJ (1993) Entrainment from a bed of particles by thermal-convection. *Earth and Planetary Science Letters* 120: 387–393.

Solomatov VS and Reese CC (2008) Grain size variations in the Earth's mantle and evolution of primordial chemical heterogeneities. *Journal of Geophysical Research* 113: B07408. http://dx.doi.org/10.1029/2007JB005319.

Solomatov VS and Stevenson DJ (1993a) Nonfractional crystallization of a terrestrial magma ocean. *Journal of Geophysical Research: Planets* 98: 5391–5406.

Solomatov VS and Stevenson DJ (1993b) Suspension in convective layers and style of differentiation of a terrestrial magma ocean. *Journal of Geophysical Research: Planets* 98: 5375–5390.

Song X and Ahrens TJ (1994) Pressure-temperature range of reactions between liquid-iron in the outer core and mantle silicates. *Geophysical Research Letters* 21: 153–156.
Song T-RA, Helmberger DV, and Grand SP (2004) Low-velocity zone atop the 410-km seismic discontinuity in the northwestern United States. *Nature* 427: 530–533.
Spence DA, Ockendon JR, Wilmott P, Turcotte DL, and Kellogg LH (1988) Convective mixing in the mantle: The role of viscosity differences. *Geophysical Journal* 95: 79–86.
Stacey FD (1992) *Physics of the Earth*, 3rd edn. Kenmore, Queensland, Australia: Brookfield Press.
Stackhouse S and Brodholt JP (2008) Elastic properties of the post-perovskite phase of Fe_2O_3 and implications for ultra-low velocity zones. *Physics of the Earth and Planetary Interiors* 170: 260–266.
Stackhouse S, Brodholt J, and Price GD (2006) Elastic anisotropy of $FeSiO_3$ end-members of the perovskite and post-perovskite phases. *Geophysical Research Letters* 33: L01304. http://dx.doi.org/10.1029/2005GL023887.
Stagno V, Ojwang DO, McCammon CA, and Frost DJ (2013) The oxidation state of the mantle and the extraction of carbon from Earth's interior. *Nature* 493: 84–88.
Stegman DR, Richards MA, and Baumgardner JR (2002) Effects of depth-dependent viscosity and plate motions on maintaining a relatively uniform mid-ocean ridge basalt reservoir in whole mantle flow. *Journal of Geophysical Research* 107. http://dx.doi.org/10.1029/2001JB000192.
Stein M and Hofmann AW (1994) Mantle plumes and episodic crustal growth. *Nature* 372: 63–68.
Steinberger B and Torsvik TH (2012) A geodynamic model of plumes from the margins of large low shear velocity provinces. *Geochemistry, Geophysics, Geosystems* 13: Q01W09.
Stern RJ and Scholl DW (2010) Yin and yang of continental crust creation and destruction by plate tectonic processes. *International Geology Review* 52: 1–31.
Stevenson DJ (1988) Infiltration, dissolution, and underplating: Rules for mixing core-mantle cocktails. *EOS (Transactions of the American Geophysical Union)* 69: 1404, Fall Meeting Supplement.
Stevenson DJ (1990) Fluid dynamics of core formation. In: Jones JH and Newman HE (eds.) *Origin of the Earth*, pp. 231–249. Oxford: Oxford University Press.
Stixrude L, de Koker N, Sun N, Mookherjee M, and Karki BB (2009) Thermodynamics of silicate liquids in the deep Earth. *Earth and Planetary Science Letters* 278: 226–232.
Stracke A (2012) Earth's heterogeneous mantle: A product of convection-driven interaction between crust and mantle. *Chemical Geology* 330–331: 274–299.
Stracke A, Hofmann AW, and Hart S (2005) FOZO, HIMU, and the rest of the mantle zoo. *Geochemistry, Geophysics, Geosystems* 6: Q05007. http://dx.doi.org/10.1029/2004GC000824.
Stracke A, Zindler A, Salters VJM, et al. (2003) Theistareykir revisited. *Geochemistry, Geophysics, Geosystems* 4(2): 8507. http://dx.doi.org/10.1029/2001GC000201.
Stuart FM, Lass-Evans S, Fitton JG, and Ellam RM (2003) High 3He/4He ratios in picritic basalts from Baffin Island and the role of a mixed reservoir in mantle plumes. *Nature* 424: 57–59.
Su WJ and Dziewonski AM (1997) Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle. *Physics of the Earth and Planetary Interiors* 100: 135–156.
Subramanian N, Kellogg LH, and Turcotte DL (2009) Statistics of advective stretching in three-dimensional incompressible flows. *Journal of Statistical Physics* 136: 926–944.
Suen CJ and Frey FA (1987) Origins of mafic and ultramafic rocks in the Ronda peridotite. *Earth and Planetary Science Letters* 85: 183–202.
Tackley PJ (1998) Three-dimensional simulations of mantle convection with a thermochemical CMB boundary layer: D″? In: Wysession ME, Knittle E, Buffett BA, and Gurnis M (eds.) *The Core-Mantle Boundary Region*, pp. 231–253. Washington, DC: American Geophysical Union.
Tackley PJ (2000) Mantle convection and plate tectonics: Towards an integrated physical and chemical theory. *Science* 288: 2002.
Tackley PJ (2002) Strong heterogeneity caused by deep mantle layering. *Geochemistry, Geophysics, Geosystems* 3. http://dx.doi.org/10.1029/2001GC000167.
Tackley PJ (2011) Living dead slabs in 3-D: The dynamics of compositionally-stratified slabs entering a "slab graveyard" above the core-mantle boundary. *Physics of the Earth and Planetary Interiors* 188: 150–162.
Tackley PJ (2012) Dynamics and evolution of the deep mantle resulting from thermal, chemical, phase and melting effects. *Earth Science Reviews* 110: 1–25.
Tackley PJ and King SD (2003) Testing the tracer ratio method for modeling active compositional fields in mantle convection simulations. *Geochemistry, Geophysics, Geosystems* 4: 8302. http://dx.doi.org/10.1029/2001GC000214.
Tackley PJ, Nakagawa T, and Hernlund JW (2007) Influence of the post-perovskite transition on thermal and thermo-chemical mantle convection. In: Hirose K, Brodholt J, Lay T, and Yuen DA (eds.) *Post-perovskite: The Last Mantle Phase Transition*. *AGU Geophysical Monograph Series*, vol. 174, pp. 229–248. Washington, DC: American Geophysical Union.
Tackley PJ, Stevenson DJ, Glatzmaier GA, and Schubert G (1994) Effects of multiple phase transitions in a 3-dimensional spherical model of convection in Earth's mantle. *Journal of Geophysical Research* 99: 15877–15901.
Tackley PJ and Xie S (2002) The thermo-chemical structure and evolution of Earth's mantle: Constraints and numerical models. *Philosophical Transactions of the Royal Society A* 360: 2593–2609.
Tackley PJ, Xie S, Nakagawa T, and Hernlund JW (2005) Numerical and laboratory studies of mantle convection: Philosophy, accomplishments and thermo-chemical structure and evolution. In: van der Hilst RD, Bass J, Matas J, and Trampert J (eds.) *AGU Geophysical Monograph on Earth's Deep Mantle: Structure, Composition and Evolution*, pp. 85–102. Washington, DC: AGU.
Tajika E and Matsui T (1992) Evolution of terrestrial proto-CO_2 atmosphere coupled with thermal history of the Earth. *Earth and Planetary Science Letters* 113: 251–266.
Tajika E and Matsui T (1993) Evolution of sea-floor spreading rate-based on Ar-40 degassing history. *Geophysical Research Letters* 20: 851–854.
Tan E and Gurnis M (2005) Metastable superplumes and mantle compressibility. *Geophysical Research Letters* 32: L20307, http://dx.doi.org/20310.21029/22005GL024190.
Tan E and Gurnis M (2007) Compressible thermochemical convection and application to lower mantle structures. *Journal of Geophysical Research* 112: B06304, http://dx.doi.org/06310.01029/02006jb004505.
Tan E, Gurnis M, and Han L (2002) Slabs in the lower mantle and their modulation of plume formation. *Geochemistry, Geophysics, Geosystems* 3: 1067, http://dx.doi.org/1010.1029/2001GC000238.
Tan E, Leng W, Zhong S, and Gurnis M (2011) On the location of plumes and lateral movement of thermo-chemical structures with high bulk modulus in the 3-D compressible mantle. *Geochemistry, Geophysics, Geosystems* 12: Q07005. http://dx.doi.org/10.1029/2011GC003665.
Tatsumi Y, Suzuki T, Ozawa H, Hirose K, Hanyu T, and Ohishi Y (2013) Accumulation of 'anti-continent' at the base of the mantle and its recycling in mantle plumes. *Geochimica et Cosmochimica Acta*. http://dx.doi.org/10.1016/j.gca.2013.11.019.
Tauzin B, Debayle E, and Wittlinger G (2010) Seismic evidence for a global low-velocity layer within the Earth/'s upper mantle. *Nature Geoscience* 3: 718–721.
Taylor SR and McLennan SM (1995) The geochemical evolution of the continental crust. *Reviews of Geophysics* 33: 241–265.
Ten AA, Podladchikov YY, Yuen DA, Larsen TB, and Malevsky AV (1998) Comparison of mixing properties in convection with the particle-line method. *Geophysical Research Letters* 25: 3205–3208.
Ten A, Yuen DA, Larsen TB, and Malevsky AV (1996) The evolution of material-surfaces in convection with variable viscosity as monitored by a characteristics-based method. *Geophysical Research Letters* 23: 2001–2004.
Ten A, Yuen DA, Podladchikov YY, Larsen TB, Pachepsky E, and Malevsky AV (1997) Fractal features in mixing of non-newtonian and newtonian mantle convection. *Earth and Planetary Science Letters* 146: 401–414.
Thorne MS, Garnero EJ, Jahnke G, Igel H, and McNamara AK (2013) Mega ultra low velocity zone and mantle flow. *Earth and Planetary Science Letters* 364: 59–67.
Thorne MS, Grand SP, and Garnero EJ (2004) Geographic correlation between hot spots and deep mantle lateral shear-wave velocity gradients. *Physics of the Earth and Planetary Interiors* 146: 47–63.
Tolstikhin IN and Hofmann AW (2005) Early crust on top of the Earth's core. *Physics of the Earth and Planetary Interiors* 148: 109–130.
Tolstikhin IN, Kramers JD, and Hofmann AW (2006) A chemical Earth model with whole mantle convection: The importance of a core-mantle boundary layer (D″) and its early formation. *Chemical Geology* 226: 79–99.
Tonks WB and Melosh HJ (1990) The physics of crystal settling and suspension in a turbulent magma ocean. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 151–174. New York: Oxford University Press.
Torsvik T, Burke K, Steinberger B, Webb SJ, and Ashwal LD (2010) Diamonds sampled by plumes from the core-mantle boundary. *Nature* 466: 352–355.
Torsvik TH, Smethurst MA, Burke K, and Steinberger B (2006) Large igneous provinces generated from the margins of the large low-velocity provinces in the deep mantle. *Geophysical Journal International* 167: 1447–1460.
Torsvik TH, Smethurst MA, Burke K, and Steinberger B (2008) Long term stability in deep mantle structure: Evidence from the ~300 Ma Skagerrak-Centered Large Igneous Province (the SCLIP). *Earth and Planetary Science Letters* 267: 444–452.
Trail D, Watson EB, and Tailby ND (2011) The oxidation state of Hadean magmas and implications for early Earth/'s atmosphere. *Nature* 480: 79–82.

Trampert J, Deschamps F, Resovsky JS, and Yuen D (2004) Probabilistic tomography maps significant chemical heterogeneities in the lower mantle. *Science* 306: 853–856.

Trieloff M, Kunz J, Clague DA, Harrison D, and Allegre CJ (2000) The nature of pristine noble gases in mantle plumes. *Science* 288: 1036–1038.

Turcotte DL, Paul D, and White WM (2001) Thorium-uranium systematics require layered mantle convection. *Journal of Geophysical Research* 106: 4265–4276.

Turcotte DL and Schubert G (1982) *Geodynamics: Applications of Continuum Physics to Geological Problems*. New York: Wiley.

Turner G (1989) The outgassing history of the Earth's atmosphere. *Journal of the Geological Society of London* 146: 147–154.

Ulvrova M, Coltice N, Ricard Y, et al. (2011) Compositional and thermal equilibration of particles, drops, and diapirs in geophysical flows. *Geochemistry, Geophysics, Geosystems* 12: Q10014.

Ulvrová M, Labrosse S, Coltice N, Råback P, and Tackley PJ (2012) Numerical modelling of convection interacting with a melting and solidification front: Application to the thermal evolution of the basal magma ocean. *Physics of the Earth and Planetary Interiors* 206–207: 51–66.

van den Berg AP, Jacobs MH, and de Jong BH (2002) Numerical models of mantle convection based on thermodynamic data for the MgOSiO2 olivine-pyroxene system. *EOS Transactions, American Geophysical Union* 83, Fall Meeting Supplement. Abstract MR72B-1041.

van der Hilst RD and Karason H (1999) Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model. *Science* 283: 1885–1888.

Van der Hilst RD, Widlyantoro S, and Engdahl ER (1997) Evidence for deep mantle circulation from global tomography. *Nature* 386: 578–584.

van der Meer DG, Spakman W, van Hinsbergen DJJ, Amaru ML, and Torsvik TH (2009) Towards absolute plate motions constrained by lower-mantle slab remnants. *Nature Geoscience* 3: 36–40.

van Keken PE and Ballentine CJ (1998) Whole-mantle versus layered mantle convection and the role of a high-viscosity lower mantle in terrestrial volatile evolution. *Earth and Planetary Science Letters* 156: 19–32.

van Keken PE and Ballentine CJ (1999) Dynamical models of mantle volatile evolution and the role of phase transitions and temperature-dependent rheology. *Journal of Geophysical Research* 104: 7137–7151.

van Keken PE, Ballentine CJ, and Hauri EH (2003) Convective mixing in the Earth's mantle. In: Carlson RW (ed.) *Treatise on Geochemistry*, pp. 471–491. Oxford: Elsevier.

van Keken PE, Ballentine CJ, and Porcelli D (2001) A dynamical investigation of the heat and helium imbalance. *Earth and Planetary Science Letters* 188: 421–434.

van Keken PE, Hacker BR, Syracuse EM, and Abers GA (2011) Subduction factory: 4. Depth-dependent flux of H_2O from subducting slabs worldwide. *Journal of Geophysical Research* 116: B01401.

van Keken PE, Karato S, and Yuen DA (1996) Rheological control of oceanic crust separation in the transition zone. *Geophysical Research Letters* 23: 1821–1824.

van Keken PE, Kiefer B, and Peacock SM (2002) High-resolution models of subduction zones: Implications for mineral hydration reactions and the transport of water into the deep mantle. *Geochemistry, Geophysics, Geosystems* 3: 1056. http://dx.doi.org/10.1029/2001GC000256.

van Keken PE, King SD, Schmeling H, Christensen UR, Neumeister D, and Doin MP (1997) A comparison of methods for the modeling of thermochemical convection. *Journal of Geophysical Research* 102: 22477–22495.

van Keken PE and Zhong S (1999) Mixing in a 3D spherical model of present-day mantle convection. *Earth and Planetary Science Letters* 171: 533–547.

van Summeren JRG, van den Berg AP, and van der Hilst RD (2009) Upwellings from a deep mantle reservoir filtered at the 660 km phase transition in thermo-chemical convection models and implications for intra-plate volcanism. *Physics of the Earth and Planetary Interiors* 172: 210–224.

van Thienen P, van den Berg AP, and Vlaar NJ (2004) Production and recycling of oceanic crust in the early Earth. *Tectonophysics* 386: 41–65.

van Thienen P, van Summeren J, van der Hilst RD, van den Berg AP, and Vlaar NJ (2005a) Numerical study of the origin and stability of chemically distinct reservoirs deep in Earth's mantle. In: van der Hilst, RD, Bass J, Matas J, and Trampert J (eds.) *The Structure, Evolution and Composition of Earth's Mantle. AGU Geophysical Monograph*, vol. 160, pp. 117–136. Washington, DC: American Geophysical Union.

van Thienen P, Vlaar NJ, and van den Berg AP (2005b) Assessment of the cooling capacity of plate tectonics and flood volcanism in the evolution of Earth, Mars and Venus. *Physics of the Earth and Planetary Interiors* 150: 287–315.

Vidale JE and Hedlin MAH (1998) Evidence for partial melt at the core-mantle boundary north of Tonga from the strong scattering of seismic waves. *Nature* 391: 682–685.

Vidale JE, Schubert G, and Earle PS (2001) Unsuccessful initial search for a midmantle chemical boundary layer with seismic arrays. *Geophysical Research Letters* 28: 859–862.

Vlaar NJ (1985) Precambrian geodynamical constraints. In: Touret JLR and Tobi AC (eds.) *The Deep Proterozoic Crust in the North Atlantic Provinces*, pp. 3–20. Dordrecht: D. Riedel Publishing Company.

Vlaar NJ (1986) Archaean global dynamics. In: *Dutch Contributions to the International Lithosphere Program; Meeting 65,* pp. 91–101.

Walker JCG, Hays PB, and Kasting JF (1981) A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *Journal of Geophysical Research* 86: 9776–9782.

Walker RJ, Morgan JW, and Horan MF (1995) 187Os enrichment in some mantle plume sources: Evidence for core-mantle interaction? *Science* 269: 819–822.

Walter MJ, Kohn SC, Araujo D, et al. (2011) Deep mantle cycling of oceanic crust: Evidence from diamonds and their mineral inclusions. *Science* 334: 54–57.

Walzer U and Hendel R (1997a) Tectonic episodicity and convective feedback mechanisms. *Physics of the Earth and Planetary Interiors* 100: 167–188.

Walzer U and Hendel R (1997b) Time-dependent thermal convection, mantle differentiation and continental-crust growth. *Geophysical Journal International* 130: 303–325.

Walzer U and Hendel R (1999) A new convection-fractionation model for the evolution of the principal geochemical reservoirs of the Earth's mantle. *Physics of the Earth and Planetary Interiors* 112: 211–256.

Walzer U and Hendel R (2008) Mantle convection and evolution with growing continents. *Journal of Geophysical Research* 113: B09405.

Wang Y and Wen L (2007) Geometry and P and S velocity structure of the "African Anomaly". *Journal of Geophysical Research* 112: B05313. http://dx.doi.org/05310.01029/02006jb004483.

Weinstein SA (1992) Induced compositional layering in a convecting fluid layer by an endothermic phase-transition. *Earth and Planetary Science Letters* 113: 23–39.

Weinstein SA (1993) Catastrophic overturn of the earths mantle driven by multiple phase-changes and internal heat-generation. *Geophysical Research Letters* 20: 101–104.

Wen L (2001) Seismic evidence for a rapidly-varying compositional anomaly at the base of the Earth's mantle beneath the Indian ocean. *Earth and Planetary Science Letters* 194: 83–95.

Wen LX (2002) An SH hybrid method and shear velocity structures in the lowermost mantle beneath the central Pacific and South Atlantic Oceans. *Journal of Geophysical Research: Solid Earth* 107. http://dx.doi.org/10.1029/2001JB000499.

Wen LX and Anderson DL (1997) Layered mantle convection – A model for geoid and topography. *Earth and Planetary Science Letters* 146: 367–377.

Wen L and Helmberger DV (1998) Ultra-low velocity zones near the core-mantle boundary from broadband PKP precursors. *Science* 279: 1701–1703.

White WM (1985) Sources of oceanic basalts: Radiogenic isotope evidence. *Earth and Planetary Science Letters* 115: 211–226.

White WM (2010) Oceanic island basalts and mantle plumes: The geochemical perspective. *Annual Review of Earth and Planetary Sciences* 38: 133–160.

Wicks JK, Jackson JM, and Sturhahn W (2010) Very low sound velocities in iron-rich (Mg, Fe)O: Implications for the core-mantle boundary region. *Geophysical Research Letters* 37: L15304.

Williams Q and Garnero EJ (1996) Seismic evidence for partial melt at the base of earths mantle. *Science* 273: 1528–1530.

Williams Q, Revenaugh J, and Garnero E (1998) A correlation between ultra-low basal velocities in the mantle and hot spots. *Science* 281: 546–549.

Wolf A (1986) Quantifying chaos with Lyapunov exponents. In: Holden AV (ed.) *Chaos*, pp. 273–290. Princeton, NJ: Princeton University Press.

Wolfe CJ, Solomon SC, Laske G, et al. (2011) Mantle P-wave velocity structure beneath the Hawaiian hotspot. *Earth and Planetary Science Letters* 303: 267–280.

Wood BJ and Blundy JD (2003) Trace element partitioning under crustal and uppermost mantle conditions: The influences of ionic radius, cation charge, pressure, and temperature. In: Carlson RW (ed.) *Treatise on Geochemistry. The Mantle and Core*, vol. 2, pp. 395–424. Amsterdam: Elsevier.

Xie S and Tackley PJ (2004a) Evolution of helium and argon isotopes in a convecting mantle. *Physics of the Earth and Planetary Interiors* 146: 417–439.

Xie S and Tackley PJ (2004b) Evolution of U-Pb and Sm-Nd systems in numerical models of mantle convection. *Journal of Geophysical Research* 109: B11204. http://dx.doi.org/10.1029/2004JB003176.

Xu W, Lithgow-Bertelloni C, Stixrude L, and Ritsema J (2008) The effect of bulk composition and temperature on mantle seismic structure. *Earth and Planetary Science Letters* 275: 70–79.

Xu-Dong L and Osher S (1996) Nonoscillatory high order accurate self-similar maximum principle satisfying shock capturing schemes I. *The SIAM Journal on Numerical Analysis* 33: 760–779.

Yabe T, Ogata Y, Takizawa K, Kawai T, Segawa A, and Sakurai K (2002) The next generation CIP as a conservative semi-Lagrangian solver for solid, liquid and gas. *Journal of Computational and Applied Mathematics* 149: 267–277.

Yamazaki D and Karato S (2001) Some mineral physics constraints on the rheology and geothermal structure of Earth's lower mantle. *American Mineralogist* 86: 385–391.

Yokochi R and Marty B (2004) A determination of the neon isotopic composition of the deep mantle. *Earth and Planetary Science Letters* 225: 77–88.

Yoshida M and Ogawa M (2004) The role of hot uprising plumes in the initiation of plate-like regime of three-dimensional mantle convection. *Geophysical Research Letters* 31: 5607.

Youngs BAR and Bercovici D (2009) Stability of a compressible hydrous melt layer above the transition zone. *Earth and Planetary Science Letters* 278: 78–86.

Yuan K and Beghein C (2013) Seismic anisotropy changes across upper mantle phase transitions. *Earth and Planetary Science Letters* 374: 132–144.

Zahnle K, Schaefer L, and Fegley B (2010) Earth's earliest atmospheres. *Cold Spring Harbor Perspectives in Biology* 2: a004895.

Zhong S (2006) Constraints on thermochemical convection of the mantle from plume heat flux, plume excess temperature, and upper mantle temperature. *Journal of Geophysical Research* 111: B04409. http://dx.doi.org/10.1029/2005JB003972.

Zhong SJ and Hager BH (2003) Entrainment of a dense layer by thermal plumes. *Geophysical Journal International* 154: 666–676.

Ziegler LB and Stegman DR (2013) Implications of a long-lived basal magma ocean in generating earth's ancient magnetic field. *Geochemistry, Geophysics, Geosystems* 14: 4735–4742.

Zindler A and Hart S (1986) Geochemical geodynamics. *Earth and Planetary Science Letters* 14: 493–571.